

From Nuisance to Novel Research Questions: Using Multilevel Models to Predict Heterogeneous Variances

Organizational Research Methods
1-47

© The Author(s) 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1094428119887434

journals.sagepub.com/home/orm



Houston F. Lester¹ , Kristin L. Cullen-Lester²,
and Ryan W. Walters³

Abstract

Constructs that reflect differences in variability are of interest to many researchers studying workplace phenomena. The aggregation methods typically used to investigate “variability-based” constructs suffer from several limitations, including the inability to include Level 1 predictors and a failure to account for uncertainty in the variability estimates. We demonstrate how mixed-effects location-scale (MELS) and heterogeneous variance models, which are direct extensions of traditional mixed-effects (or multilevel) models, can be used to test mean (location)- and variability (scale)-related hypotheses simultaneously. The aims of this article are to demonstrate (a) how the MELS and heterogeneous variance models can be estimated with both nested cross-sectional and longitudinal data to answer novel research questions about constructs of interest to organizational researchers, (b) how a Bayesian approach allows for the inclusion of random intercepts and slopes when predicting both variability and mean levels, and finally (c) how researchers can use a multilevel approach to predict between-group heterogeneous variances. In doing so, this article highlights the added value of viewing variability as more than a statistical nuisance in organizational research.

Keywords

multilevel models, mixed-effects location-scale models, heterogeneous variance models, variability-related hypotheses, Bayesian

Organizational researchers commonly use the amount (e.g., mean) of a variable as both a predictor and an outcome of substantive interest, while examining variability for instrumental purposes. For example, researchers using direct consensus or referent shift approaches (Chan, 1998) calculate

¹Baylor College of Medicine, Houston TX, USA

²University of Houston Bauer College of Business, Houston, TX, USA

³School of Medicine, Creighton University, Omaha, NE, USA

Corresponding Author:

Houston F. Lester, Baylor College of Medicine, 2450 Holcombe Blvd., Suite 01Y, Houston, TX 77021, USA.

Email: hfl000151@gmail.com

agreement among members of a group to justify the aggregation of a Level 1 variable (James, Demaree, & Wolf, 1984, 1993). Similarly, before proceeding with their analysis, researchers using general linear models assess whether residuals follow the assumed normal distribution with a homogeneous variance (e.g., Kutner, Nachtsheim, Neter, & Li, 2004). In such cases, researchers may view variability as a nuisance that interferes with obtaining appropriate estimates and inferences regarding the mean. However, theoretically meaningful interpretations of variability exist in many areas of organizational research (e.g., personality consistency, performance variability, leader-member exchange [LMX] differentiation, strategic consensus, pay dispersion) and researchers need to utilize methods designed for studying variability to advance the understanding of these workplace phenomena.

Currently, it is common for researchers publishing in top organizational journals to use single-level aggregation methods in analyses predicting variability (e.g., calculating the within-group standard deviation and using it as an outcome in analyses). However, we caution against this common practice due to important limitations, including a failure to account for uncertainty in the variability estimates and the inability to appropriately incorporate Level 1 predictors in analyses. Instead, we propose that the study of variability is ripe for multilevel investigation and explain how multilevel approaches enable researchers to address new research questions that cannot be answered using aggregation and single-level analyses. Specifically, we demonstrate how models that are direct extensions of traditional mixed-effects (or multilevel) models may be used to test mean (location)- and variability (scale)-related hypotheses simultaneously. The most general of these models, **mixed-effects location-scale (MELS) models**, allow for **random effects** on the location side and the scale side of the model, while the more restricted **heterogeneous variance models** allow for random effects on the location side of the model and **fixed effects** on the scale side of the model (see the glossary in Appendix A for definitions of all bold terms).

We are not the first to discuss MELS or heterogeneous variance models. MELS models are a relatively recent development in the long history of different approaches to modeling heterogeneous variances (e.g., Aitkin, 1987; Bryk & Raudenbush, 1988; Culpepper, 2010; Goldstein, 2011; Harvey, 1976; Leonard, 1975; Lindley, 1971; Pinheiro & Bates, 2000; Raudenbush, 1988). Researchers have discussed the development of these models in statistical journals and texts (e.g., Goldstein, 2011; Hedeker, Mermelstein, & Demirtas, 2008; Lin, Mermelstein, & Hedeker, 2018a, 2018b; Walters, Hoffman, & Templin, 2018) and applied these models (most often) in fields where the collection of intensive longitudinal data is more common (e.g., medicine; Pugach, Hedeker, Richmond, Sokolovsky, & Mermelstein, 2014). For example, Watts, Walters, Hoffman, and Templin (2016) examined whether time-invariant (Level 2 predictors, e.g., gender, age, Alzheimer's disease status) as well as time-varying predictors (Level 1 predictors, e.g., day monitor worn) were associated with individual differences in mean level (location side) as well as intraindividual variability (scale side) of physical activity. Although less common, researchers also apply these models to nested cross-sectional data. For example, this approach has been used in an educational context to assess whether the variability in academic achievement within a school is affected by children's socioeconomic status (SES; Leckie, French, Charlton, & Browne, 2014). By using a Bayesian approach to include a random effect for SES on the scale side of the model, Leckie et al. (2014) demonstrated that the effect of SES on variability in academic achievement within schools varied by private versus public sector (i.e., a cross-level interaction). Culpepper (2010) employed a similar approach to reveal that whites and females demonstrate more predictable academic performance than their male or racial/ethnic minority counterparts.

The first use of these models within the organizational sciences occurred recently, with an important and innovative application to the study of consensus emergence by Lang, Bliese, and de Voogt (2018). With this exception, we suspect MELS and heterogeneous variance models are underutilized in the organizational research literature for at least two reasons. First, it is relatively

rare for organizational researchers to have intensive longitudinal data. Second, most articles have been published in statistical and medical journals that organizational scholars tend not to read, and these articles may not be accessible to all readers (e.g., they often assume a depth of mathematical training beyond the level typically provided in graduate programs training organizational researchers).

The objective of this article is to broaden the use of multilevel approaches for studying variability in the organizational sciences. We begin by identifying examples of **variance-based constructs** that are of interest to organizational researchers, and throughout the article we illustrate the types of research questions that multilevel approaches for predicting heterogeneous variances can answer, including questions that cannot be asked or adequately addressed using typical methodological practices (i.e., aggregation and single-level analyses). Then, using simulated data, we first demonstrate the estimation of a MELS model predicting heterogeneous within-group variance for a nested cross-sectional design.¹ Second, we demonstrate the estimation of a heterogeneous variance model for predicting between-group variances (i.e., intercept and slope) for a longitudinal design using a multigroup approach. In addition to predicting heterogeneous between-group variances for the intercept and slope, this model also allows for a heterogeneous covariance between these two random effects.

The flexibility of modeling the scale side with random effects does have computational costs, especially as the number of random effects increases (Asparouhov & Muthén, 2016). Researchers will likely need to use a Bayesian analysis to estimate more complex MELS models (e.g., those with both random intercepts and slopes on the location and scale sides of the model as is done in the nested cross-sectional example). Throughout the article, as we demonstrate this approach, we aim to strike a balance between providing the technical detail and practical guidance researchers need to use these models in their work. To that end, we use an open-source statistical programming language to conduct the analyses, provide all code used to estimate the models in our examples, and offer detailed interpretations of the model output. Finally, we discuss additional practical considerations for researchers applying multilevel models to predict heterogeneous variance.

Review of the Organizational Literature

In this section, we present the results of a literature review examining the study of theoretically meaningful interpretations of variability in organizational research (i.e., variability-based constructs). We start our review in the year in which Hedeker and colleagues (2008) demonstrated how to estimate the random intercept MELS model in *Biometrics* using readily available software, thereby making this method more readily available to researchers. We identified articles by searching for the terms *variability*, *dispersion*, *consensus*, and *consistency* in article titles, keywords, and abstracts of the following top journals: *Academy of Management Journal*, *Administrative Science Quarterly*, *Journal of Applied Psychology*, *Journal of Management Studies*, *Journal of Management*, *Management Science*, *Organization Science*, *Personnel Psychology*, and *Strategic Management Journal*. We identified over 60 articles in which variability-based constructs were discussed (conceptually) or examined as the substantive focus of the article (e.g., as predictor, dependent variable, moderator). While not exhaustive (e.g., researchers may use other terms to refer to variance-based constructs), this search allows us to illustrate the interest in and methods used to examine variability-based constructs in organizational research. We found that organizational researchers have not widely adopted MELS models or similar approaches.

Examples of Variability-Based Constructs in Organizational Research

Across the micro-macro organizational research continuum, we identified many examples of variability-based constructs. These constructs reflect between-person, -team, or -organization

differences in variability across time/situations (e.g., stability of preferences, attitudes, behaviors) as well as variability in the characteristics, perceptions, or behaviors of lower-level entities (e.g., people, but also teams, business units) that compose a higher-level entity (e.g., team, occupation, organization, or industry). Table 1 provides a description of these constructs.

Many variability-based constructs have broad relevance across levels and areas of focus. For example, researchers have examined performance variability for firms over time (e.g., Wales, Patel, & Lumpkin, 2013; Wang & Choi, 2013), within an industry (e.g., Balasubramanian & Lieberman, 2010; Lenox, Rockart, & Lewin, 2010), across organizational subsidiaries (Figueiredo, 2011), across people who occupy the same type of job (Bidwell & Keller, 2014), and for employees over time (Barnes, Reb, & Ang, 2012; Reb & Greguras, 2010). Similarly, pay dispersion has been discussed in reference to labor markets (Cobb & Stevens, 2017), organizations as a whole (Carnahan, Agarwal, & Campbell, 2012), and perhaps most commonly, top management teams (e.g., Chin & Semadeni, 2017; Fredrickson, Davis-Blake, & Sanders, 2010; Jaskiewicz, Block, Miller, & Combs, 2017; Lim, 2018).

Several constructs also reflect variability in characteristics of group members (e.g., personality heterogeneity) or the extent to which there is agreement among members in the group (e.g., justice climate, efficacy dispersion). For example, strategic consensus regarding the firm's priorities is commonly assessed for managers in the top team, but also at different levels and in different parts of the organization (Kellermanns, Walter, Lechner, & Floyd, 2005). The variability of supervisor behavior across time (e.g., justice variability; Matta, Scott, Colquitt, Koopman, & Passantino, 2017) and when interacting with different members of the workgroup (LMX differentiation; Gooty & Yammarino, 2016; Henderson, Liden, Glibkowski, & Chaudhry, 2009) has also been a topic of study.

It is important to note that most prior research has either conceptually discussed the nature of variance-based constructs or aimed to establish relationships between the construct and important workplace outcomes. For example, Ployhart, Weekley, and Baughman (2006) found that human capital emergence, conceptualized as consisting of both personality level (aggregate mean personality) and personality homogeneity (aggregate standard deviation of personality) within jobs and within organizations, predicted employee job satisfaction and performance. Further, research focusing on within-person personality consistency has found that the relationship between the level of personality traits and job performance is stronger when people are less variable in their personality expression across time/situations (Dalal et al., 2015). Other research has linked variability in the quality of relationships between a supervisor and his or her subordinates (i.e., LMX differentiation) to job satisfaction, performance, and other outcomes (e.g., Erdogan & Bauer, 2010; Schyns, 2006). In the literature on workplace groups and teams, researchers have proposed that team efficacy dispersion will predict team effectiveness above and beyond the average level of efficacy within the team (DeRue, Hollenbeck, Ilgen, & Feltz, 2010). At the department or organization level, researchers have found a stronger relationship between climate level (i.e., mean rating of perceptions of the workplace) and outcomes of interest (e.g., such as work satisfaction and organizational commitment) when employees view the climate similarly (i.e., there is a strong climate in place; González-Romá, Peiró, & Tordera, 2002). Finally, at the firm level, firms with higher trading variability have been found to have lower expected returns (Chordia, Subrahmanyam, & Anshuman, 2001).

These examples are part of the mounting evidence suggesting that variability (in addition to the average amount) of workplace phenomena predicts important workplace outcomes. Thus, there is a need to understand the antecedents of variability-based constructs. In the next section, we describe the smaller subset of studies (13) that have begun to examine these antecedents (see Table 2), including the limitations of approaches that have been applied previously. We explain how MELS

Table 1. Examples of Variability-Based Constructs in Organizational Research.

Differences in the Heterogeneity of the Entities Composing Teams, Business Units, Occupations, Organizations, or Industries	
Strategic consensus	The shared understanding of strategic priorities (i.e., agreement) among managers at the top, middle, or operating levels of the organization (Kellermanns et al., 2005).
Investor sentiment agreement	Heterogeneous perceptions of aggregate economic growth among investors (Cen, Lu, & Yang, 2013).
Competency rating consensus	Agreement among job incumbents on the ratings of job competencies (Lievens, Sanchez, Bartram, & Brown, 2010).
Justice climate strength	Variability in team members' justice ratings (Roberson & Williamson, 2012).
Group consensus	Shared perceptions and feelings or climates within groups (e.g., job satisfaction; Lang et al., 2018). Applicable to a variety of cognitive, motivational, and affective <i>emergent</i> states among group members (Ilgen, Hollenbeck, Johnson, & Jundt, 2005; Marks, Mathieu, & Zaccaro, 2001).
Leader-member exchange differentiation	Differentiated exchanges leaders create with subordinates within the same workgroup (Erdogan & Bauer, 2010; Gooty & Yammarino, 2016; Henderson et al., 2009; Liden, Erdogan, Wayne, & Sparrowe, 2006).
Occupational heterogeneity	Variability in knowledge, skills, abilities, and other competencies of people in the same occupation (Ployhart, Weekley, & Baughman, 2006; Sitzmann, Ployhardt, & Kim, 2019).
Team personality/need dispersion	Variability in team members personalities (Gonzalez-Mulé, DeGeest, McCormick, Seong, & Brown, 2014) or the need for power, achievement, and affiliation (Chun & Choi, 2014).
Efficacy dispersion	Within-team variability in perceptions of team efficacy (DeRue, Hollenbeck, Ilgen, & Feltz, 2010).
Pay dispersion	Within-team variability in compensation (e.g., among top managers; Carnahan, Agarwal, & Campbell, 2012; Chin & Semadeni, 2017; Fredrickson, Davis-Blake, & Sanders, 2010; Jaskiewicz et al., 2017; Lim, 2018; Messersmith, Guthrie, Ji, & Lee, 2011; Trevor, Reilly, & Gerhart, 2012).
Decision-making heterogeneity	Variability of group members' decision-making preferences (Melkonyan & Safra, 2016).
Client heterogeneity	Variability in the characteristics of clients served (e.g., patients physical and psychological characteristics; Chowdhury & Endres, 2010).
Implementation variability	Variability in the implementation of organizational changes, human resource management systems, or other policies and procedures across organizational subunits (e.g., the implementation of high performance work systems; Pak & Kim, 2018).
Innovation heterogeneity	Within-organization variability in innovation output (e.g., variability in the forward citations of patents submitted by an organization; Patel, Kohtamäki, Parida, and Wincent (2015); variability in the innovativeness of subsidiaries, Figueiredo, 2011).
Performance heterogeneity	Within-industry variability in firms' performance or profitability (Balasubramanian & Lieberman, 2010; Lenox, Rockart, & Lewin, 2010) or within-occupation variability in individuals' job performance (Bidwell & Keller, 2014).
Between-Person, -Team, or -Organization Differences in Variability Over Time or Across Situations	
Decision-making consistency	Within-person variability in decisions made in the face of similar choice occasions (Melkonyan & Safra, 2016).
Transfer of training consistency	Within-person variability in the application of newly acquired knowledge and skills to the job context (Huang, Ford, & Ryan, 2017).

(continued)

Table 1. (continued)

Personality consistency	Within-person variability in expression of personality traits or other individual differences across situations (i.e., more general intraindividual variability, Dalal et al., 2015; e.g., conscientiousness, Minbashian, Wood, & Beckmann, 2010; sociability and dutifulness, Lievens et al., 2018; mastery goal orientation, Huang et al., 2017).
Emotional labor consistency	Within-person variability in the use of surface acting and deep acting across situations (over time; Gabriel & Diefendorff, 2015; Scott, Barnes, & Wagner, 2012).
Job satisfaction variability	Within-person variability in person-organization and person-job fit perceptions (Gabriel, Diefendorff, Chandler, Moran, & Greguras, 2014).
Justice variability	Within-person stability of fairness over time (Matta, Scott, Colquitt, Koopman, & Passantino, 2017).
Performance variability	Within-person variability of performance over time (Barnes, Reb, & Ang, 2012; Dalal, Bhawe, & Fiset, 2014; Minbashian & Luppino, 2014; Reb & Greguras, 2010); within-business unit (Guo, 2017) or within-firm (Wales, Patel, & Lumpkin, 2013) variability of performance over time; within-firm variability of corporate social performance over time (Wang & Choi, 2013), business acquisition rate or success (Kim, Finkelstein, & Haleblan, 2015; Laamanen & Keil, 2008), and provision of services to a customer over time (Sriram, Chintagunta, & Manchanda, 2015).
Turnover volatility	Within-firm variability of employee turnover over time (Hausknecht & Holwerda, 2013).

and heterogeneous variance models address these limitations and expand the potential questions researchers may address when predicting variability as a substantive construct of interest.

Use and Limitations of Aggregation Approaches for Predicting Differences in Variability

Our review found that researchers commonly calculate and use aggregate measures of variability as an outcome in single-level regressions or similar analyses. This approach has important limitations, which researchers can overcome by using multilevel approaches for predicting differences in variability.

First, this approach allows researchers to utilize only Level 2 predictors because Level 2 outcome variables (e.g., the standard deviation estimate most commonly used) do not have any Level 1 variability (Leckie et al., 2014). Thus, researchers cannot incorporate Level 1 variables as predictors of variability-based constructs and instead often create Level 2 versions of Level 1 predictors and utilize them in a single-level regression (Hofmann, 2002). This approach loses information that is likely of interest to researchers. In the sections that follow, we demonstrate how the ability to incorporate Level 1 predictors offers new avenues for organizational researchers.

Second, the aggregation approach does not account for uncertainty in the variability estimates. The standard deviation or variance of a sample is random. Therefore, there is uncertainty associated with that statistic. Further, the standard deviation of the variance (i.e., $2\sigma^2/\sqrt{n}$) is potentially much larger than the standard deviation of the mean (i.e., the standard error of the mean: i.e., σ/\sqrt{n} ; Pawitan, 2001, p. 59). This uncertainty is important because aggregation approaches treat all variability outcomes as equally precise regardless of the number of measurements per Level 2 entity (e.g., person or group), resulting in an underestimation of sources of variation and therefore inflated Type I error rates. In contrast, MELS and heterogeneous variance models appropriately weight each Level 2 estimate according to the amount of data at Level 1, thereby resulting in more appropriate variance estimates and inferences.

Table 2. Prior Organizational Research Predicting Variability-Based Constructs.

Citation	Research Description	Analysis	Type of Variability
Fredrickson, Davis-Blake, and Sanders (2010)	Examined if the percentage of the team on the board, variation in team members stock ownership, and members' average tenure predicted variability of pay among members of the top management team.	Fixed effect regression was used to predict the coefficient of variation for pay.	Variability within a team
Lenox, Rockart, and Lewin (2010)	Examined if the extent of interdependencies among firms predicted variability of firm profits within an industry.	Mixed (multilevel) model, where the standard deviation of Tobin's q was used as the outcome.	Variability of firm profits within an industry
Lievens et al. (2010)	Examined whether aspects of the occupation (e.g., complexity, context, work activities) predicted the agreement (less variability) in job incumbent ratings of the job competencies.	Generalizability coefficient (representing agreement among raters) predicted by nature of the occupation in a regression.	Variability in job incumbents' ratings
González-Benito, Aguinis, Boyd, and Suárez-González (2012)	Examined if strategic consensus regarding competitive methods predicted strategic variability regarding managers' perceptions of strategic priorities (i.e., strategic consensus regarding objectives).	Distance measure calculated for each firm and used subsequently in a structural equation model.	Variability in managers' perceptions within a firm
Roberson and Williamson (2012)	Examined if team network density and self-monitoring behavior predicted variability in team members' perceptions of justice (i.e., justice climate strength).	Calculated standard deviation of justice climate perceptions and used as a dependent variable in a regression.	Variability within a team
Scott, Barnes, and Wagner (2012)	Examined if self-monitoring predicted emotional labor variability (for both surface and deep acting).	Calculated standard deviations for emotional labor variables over two weeks for each person and used that measure to calculate partial correlations with self-monitoring.	A person's variability over time
Patel et al. (2015)	Examined if entrepreneurial orientation and absorptive capacity predicted variability in forward citations of a patent approved and/or filed (i.e., innovation) within a firm.	Calculated the standard deviation of the forward citations (after adjusting for industry, etc.) to reflect variability and predicted within a structural equation modeling framework.	Variability within a firm
Chin and Semadeni (2017)	Examined if CEO liberalism predicted the variability of pay among members of the top management team (not including the CEO).	Generalized estimating equations (GEE) where the inverse coefficient of variation for pay was used as the outcome.	Variability within a team
Jaskiewicz et al. (2017)	Examined if the type of firm ownership (i.e., founder, family, or later generation owner)	Used aggregate measure of variation ($100 \times$ the coefficient of variation for top four	Variability within a team

(continued)

Table 2. (continued)

Citation	Research Description	Analysis	Type of Variability
Matta, Scott, Colquitt, Koopman, and Passantino (2017)	predicted the variability of pay among members of the top management team (not including the CEO). Examined if supervisors' self-control predicted variability in overall justice perceived by subordinates.	managers) for total compensation of four top non-CEO managers as the outcome in a regression. Used multilevel path analysis but included the standard deviation of justice perceptions across time as a Level 2 variable.	A person's variability over time
Lang, Bliese, and de Voogt (2018)	Examined whether collective efficacy predicted the trajectory of consensus in job satisfaction over time (Study 1) and whether leaders show different patterns of consensus than other group members (Study 2).	Extended a standard multilevel methodology by examining residual variances within a growth model to account for change in group consensus and included group and individual-level predictors to explain the emergence of greater within-group agreement over time.	Variability within a team over time
Lievens et al. (2018)	Examined if self-rated self-monitoring and functional flexibility predicted variability in respondents' answers to a situational judgment test examining the traits of sociability and dutifulness.	Used an IRT variance partitioning approach. In the second study, used the within-person standard deviation of personality states as the outcome and predicted it in a regression using intraindividual variability on SJT among other things as a predictor. In a footnote mentioned using multilevel modeling approaches in supplementary analyses, but a description of the specific method and results were not reported.	A person's variability across situations
Sitzmann, Ployhardt, and Kim (2018)	Examined whether aspects of occupational strength (e.g., task significance) predicted the variability of incumbents' personalities within occupations.	Used an aggregate measure of heterogeneity (mean Euclidean distance, replicated with standard deviation) in a structural equation model.	Variability within an occupation

Explaining Sources of Variance Versus Predicting Heterogeneous Variances

The use of multilevel models for predicting the location (mean) is common across many areas in the organizational sciences. These models are often “mixed-effects” models because they include both fixed and random effects. For example, a random intercept multilevel model (i.e., the simplest multilevel model) is a “mixed” model that allows each Level 2 unit to have its own predicted mean (e.g., Bliese & Ployhart, 2002; Pinheiro & Bates, 2000). Adding fixed or random effects enable researchers to investigate change in mean-level based on either Level 1 or Level 2 characteristics. For example, it is common for researchers to employ growth

		Location side (mean)	Scale side (variability)
Regression Fixed effects predicting location and single residual variance on scale side	(1)	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W})$ $\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W})$	$\sigma_{e_{ij}} \text{ or}$ $\log(\sigma_{e_{ij}}) = \tau_{00}$
Multilevel Model Fixed and random effects predicting location and single residual variance on scale side	(2)	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}) + u_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}) + u_{1j}$	$\sigma_{e_{ij}} \text{ or}$ $\log(\sigma_{e_{ij}}) = \tau_{00}$
Heterogeneous Variance Model Fixed and random effects predicting location and fixed effects predicting scale	(3)	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}) + u_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}) + u_{1j}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j} + \alpha_{1j}(X_{ij} - \bar{X}_j)$ $\alpha_{0j} = \tau_{00} + \tau_{01}(W_j - \bar{W})$ $\alpha_{1j} = \tau_{10} + \tau_{11}(W_j - \bar{W})$
Mixed-effects Location-Scale Model Fixed and random effects predicting location and fixed and random effects predicting scale	(4)	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$ $\beta_{0j} = \gamma_{00} + \gamma_{01}(W_j - \bar{W}) + u_{0j}$ $\beta_{1j} = \gamma_{10} + \gamma_{11}(W_j - \bar{W}) + u_{1j}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j} + \alpha_{1j}(X_{ij} - \bar{X}_j)$ $\alpha_{0j} = \tau_{00} + \tau_{01}(W_j - \bar{W}) + v_{0j}$ $\alpha_{1j} = \tau_{10} + \tau_{11}(W_j - \bar{W}) + v_{1j}$

Figure 1. Comparison of regression, multilevel, heterogeneous variance, and mixed-effects location-scale model equations. The location or mean side of the regression equation is presented in a multilevel form, although it contains no additional random effects and is therefore not a multilevel model. This notation is used to facilitate the comparison of the equations across the various model specifications. Note that there is no Level 1 residual term on the scale side of the model (i.e., no e , which is similar to a generalized linear model [e.g., logistic regression] that also lacks a Level 1 residual). In MELS and heterogeneous variance models, the log of the variance (or standard deviation) is used as the outcome to ensure that the predicted variance is nonnegative.

models, such as those discussed by Bliese and Ployhart (2002), with data gathered using experience sampling methods in which participants respond to prompts on many occasions (Bolger & Laurenceau, 2013). When researchers use these models, the first step involves variance partitioning (i.e., quantify the amount of variability that exists at higher and lower levels of analysis) and then subsequently explaining some portion of the variance at each level (by predicting differential mean levels with variables of interest). The variance partitioning in these models is the same for everyone in the sample (e.g., one random intercept variance, one random slope variance, and one residual variance).

The characteristic that differentiates MELS and heterogeneous variance models from traditional multilevel models is the explicit modeling of variability (i.e., heterogeneous residual variances). The location side (mean portion) of the model is the same and can be expressed using the notation popularized by Raudenbush and Bryk (2002). The equation on the scale side (i.e., residual variability portion) takes a similar form but uses a different notation to make it easier to keep track of which side of the model the parameter is located. Thus, typical growth models (and other traditional multilevel models) are a special case of MELS and heterogeneous variance models where the scale side of the model contains only the fixed intercept, τ_{00} . To clarify the difference between these types of models, see example equations for traditional regression, multilevel, heterogeneous variance, and MELS models provided in Figure 1 (compare Equation 2 for the traditional multilevel model to Equation 3 for the heterogeneous variance model and Equation 4 for the MELS model). Figures 2 and 3 further demonstrate this conceptual distinction between explaining sources of variance and

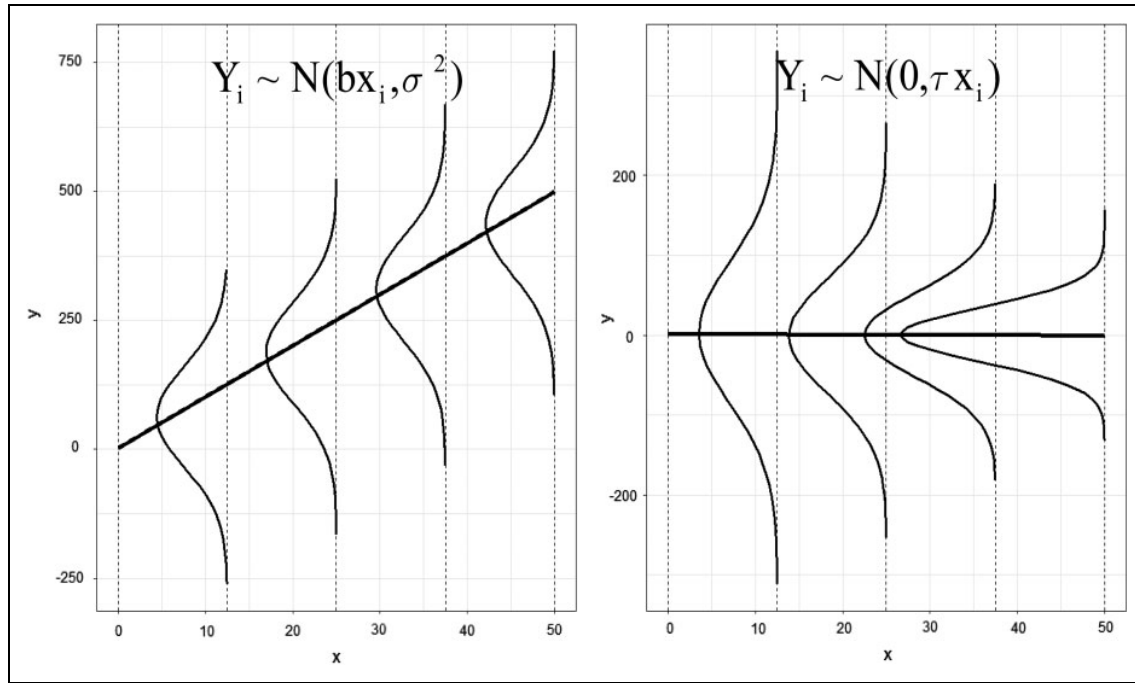


Figure 2. Regression plot illustrating regression assumptions. The left panel contains a typical regression with a positive slope and a homogenous variance. $Y_i \sim N(bx_i, \sigma^2)$ is the common notation used to indicate that Y_i is normally distributed with a mean that depends upon the individual's x_i value and a variance that is constant (σ^2). The right panel displays a heterogeneous variance model where the mean does not change over time depending upon x . The $Y_i \sim N(0, \tau x_i)$ notation indicates that the mean is zero for everyone in the sample and the variability depends upon the x_i value.

predicting heterogeneous variances (i.e., differences in variability). The latter may be used to operationalize and predict variance-based constructs in the organizational literature.

Multilevel Models for Predicting Heterogeneous Variances

Above we discussed the conceptual differences between traditional multilevel models and multilevel models for predicting heterogeneous variances. In this section, we describe the notation presented in Figure 1 for the MELS model. In these descriptions, we use individuals nested within groups to make the interpretations more concrete. However, these interpretations would be similar for different nested cross-sectional and longitudinal designs. Also, please note that the interpretation for each coefficient is the same when the coefficients appear in the other more restricted models presented in Figure 1.

First, the location (mean) side of the model in Equation 4 (see Figure 1) is just like any other multilevel model: Y_{ij} represents the outcome for individual i nested within group j , X_{ij} represents a Level 1 predictor variable (centered at the group mean, \bar{X}_j), W_j represents a Level 2 predictor (centered at the grand mean, \bar{W}). Please note that the predictors do not have to be the same for the two sides of the model. As usual, β_{0j} and β_{1j} are placeholders for the intercepts and slopes, respectively, and are defined in terms of fixed and random effects (i.e., γ 's and u 's). For interpretation purposes, individual-level (Level 1) predictors predict differential mean outcome values between individuals (that could vary across groups depending on random slopes), whereas group-level (Level 2) variables predict differential mean outcomes between groups—each unit increase in a predictor with a positive fixed effect will result in an increase in the average amount of the outcome.

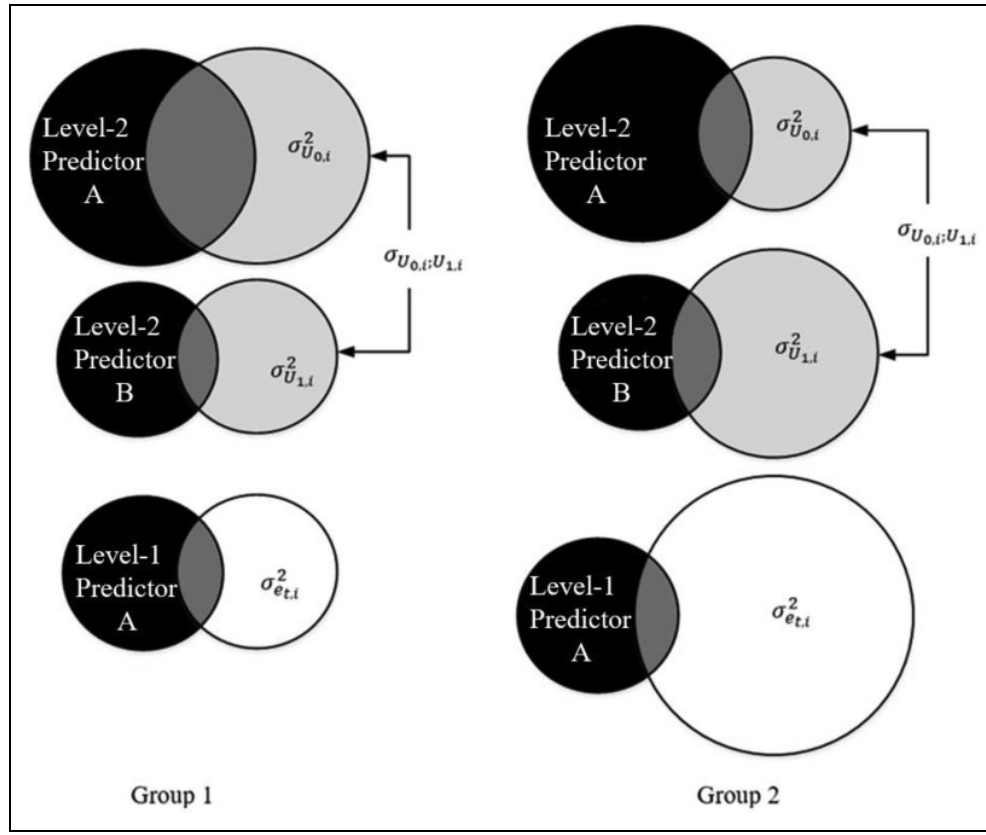


Figure 3. Conceptual figure depicting predicted variances for two different groups in a MELS model. The black circles represent three different predictors (two at Level 2 and one at Level 1). The other circles represent the variance partitioning between Level 1 (white) and Level 2 (gray). The overlap between circles represents the variance explained. The size of the circle reflects group specific residual variance estimates. Group 2 has less random intercept variance (top circle) but more random slope (middle circle) and Level 1 (bottom circle) variance as compared to Group 1. In a traditional random slope multilevel model there are only three variance components for everyone in the entire sample. In other words, the size of the circles would be the same for all groups. Although not depicted here, the same principles apply to individuals.

Second, the scale (variability) side of the MELS model is a log-linear model in which the log of the residual variance or standard deviation (we use the standard deviation parameterization) is predicted by a linear function using predictors of interest. The log-linear model is used to ensure that the predicted standard deviation is positive. As shown on the scale side of Equation 4, each group (j) and individual within that group (i) has its *own* residual standard deviation estimate, $\sigma_{e_{ij}}$ (i.e., heterogeneous). The interpretation of the model effects on the scale side is similar to the interpretation of the model effects on the location side (assuming the interpretation is made on the log standard deviation scale). Here, $\sigma_{e_{ij}}$ denotes the residual standard deviation for a particular individual in a particular group, α_{0j} and α_{1j} are placeholders for the intercepts and slopes respectively that are defined by the fixed (i.e., τ 's) and random (i.e., v 's) effects. For interpretation purposes, individual-level (Level 1) predictors predict differential within-group variability between individuals, whereas group-level (Level 2) predictors predict differential variability between groups—each unit increase in a predictor with a positive fixed effect will result in increased (log) variability indicating more variable, or less consistent, group members.

Finally, this model generally assumes all random effects v 's and u 's are normally distributed with means of 0 and random effect standard deviations σ_{u_q} and σ_{v_q} , respectively. Further, covariances

between all location and scale random effects can be estimated (e.g., σ_{u_q, v_s}). The examination of covariances allows researchers to answer questions such as do groups that have higher means also tend to have smaller within-group residual standard deviations. Equation 5 contains the assumptions regarding the distribution of the random effects (i.e., multivariate normal with a $\mathbf{0}$ mean column vector, and an unstructured covariance matrix that allows all possible covariances between the random effects). Allowing the random effects to covary across sides of the model is a unique aspect of the MELS model that cannot be estimated using the aggregation approach described previously. We provide a full interpretation of these effects in the supplemental material. Please note that the parameterization utilizes standard deviations and correlations as opposed to variances and covariances.

$$\begin{bmatrix} u_{0j} \\ u_{1j} \\ v_{0j} \\ v_{1j} \end{bmatrix} \sim MVNormal \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0, u_1} & \sigma_{u_0, v_0} & \sigma_{u_0, v_1} \\ \sigma_{u_1, u_0} & \sigma_{u_1}^2 & \sigma_{u_1, v_0} & \sigma_{u_1, v_1} \\ \sigma_{v_0, u_0} & \sigma_{v_0, u_1} & \sigma_{v_0}^2 & \sigma_{v_0, v_1} \\ \sigma_{v_1, u_0} & \sigma_{v_1, u_1} & \sigma_{v_1, v_0} & \sigma_{v_1}^2 \end{bmatrix} \right) \quad (5)$$

If the scale side of the model contains fixed, but not random, effects of predictors (i.e., no v_{0j} or v_{1j}), then the model is considered to be a heterogeneous variance model. As described previously, heterogeneous variance models are a restricted version of the MELS model (compare Equation 3 to Equation 4 in Figure 1). This comparison is similar to how linear regression is a restricted version of multilevel models. Heterogeneous variance models, by their inclusion of only fixed effects on the scale side, predict variances that differ for systematic reasons (e.g., males are more variable than females). Recently, Lang et al. (2018) demonstrated the use of an extended multilevel model for heterogeneous variances to predict the emergence of within-group consensus over time. In their article, they estimated several models with random slopes on the location side and the fixed effect of time on the scale side. Some of these models included additional moderators of the effect of time on the scale side (e.g., leadership status and group readiness).

We build on this work by Lang and colleagues (2018) by demonstrating how random slopes may also be included on the scale side of the model. This extension is important because, in a different context, Leckie et al. (2014) and Walters et al. (2018) found that ignoring random effects on the scale side can lead to inflated Type I error rates when predicting heterogeneous within-group variances, specifically when testing hypotheses regarding Level 2 variables. Thus, heterogeneous variance models are susceptible to inflated Type I error rates if additional random effects are erroneously omitted. We want to emphasize that we are not saying the application of heterogeneous variance models is wrong, but rather using these models assumes a systematically varying cross-level interaction (i.e., the variances are heterogeneous due to time as well as other fixed effect interaction terms; Davidian & Giltinan, 1995; Hoffman, 2007; Raudenbush & Bryk, 2002). It is possible to test this assumption by examining the scale-model random effects in the more general MELS models.

A Bayesian Approach

Several years ago, Kruschke, Aguinis, and Joo (2012) outlined a case for the need to apply Bayesian methods in the organizational sciences (see Appendix B for a brief introduction to Bayesian statistics). Kruschke et al. recommend five steps researchers should follow when presenting the findings of a Bayesian analysis. We use these step-by-step guidelines as we make a case for the use of Bayesian methods to estimate MELS and heterogeneous variance models and present the results from our illustrative examples.

First, it is important for researchers to motivate the use of Bayesian methods. While researchers may estimate simple versions of multilevel models predicting heterogeneous variances using

maximum likelihood estimation, convergence issues are likely to occur as the number of random effects increase on the location and scale sides of the model (Asparouhov & Muthén, 2016). Thus, we use Bayesian analysis to “open the door to extensive new realms of modeling possibilities that were previously inaccessible” (Kruschke et al., 2012, p. 723). In other words, Bayesian approaches allow researchers to estimate complex models that would be extremely difficult if not impossible to estimate using frequentist approaches (e.g., MELS models that have random intercepts and slopes on both the location and scale sides of the model).

Second, researchers should describe the model and its parameters in detail. We provide this information for both of our illustrative examples in Appendices D and E as well as in the supplemental material. Third, researchers should justify the use of the prior in their analysis. The priors we used were intended to be uninformed (i.e., should not affect the **posterior distribution**). Thus, for each example, we justify the selection of our prior by performing a sensitivity analysis to demonstrate that the use of different priors did not affect the results.

Fourth, Kruschke et al. (2012) encouraged researchers to describe the Markov chain Monte Carlo (MCMC) process in detail (see Appendix C). The models are estimated using Stan (Carpenter et al., 2017) through the brms R package (Bürkner, 2017). Stan is a powerful probabilistic programming language that allows researchers to estimate many different types of complex models. The Stan programming language is similar to other open-source Bayesian software like OpenBUGS (Spiegelhalter, Thomas, Best, & Lunn, 2007) or JAGS (Plummer, 2013). However, working with the Stan code directly is more difficult than working with these other software programs or typical R packages. The brms package was written to reduce this burden (Bürkner, 2017) and may be used by researchers to generate efficient Stan code. The brms package requests that you provide code that is similar to typical multilevel model code in R (e.g., see the lme4 R package; Bates, Maechler, Bolker, & Walker, 2015). The brms code is then translated into the Stan language (i.e., brms creates the Stan code as well as the Stan data file).

Estimating a Bayesian MELS or heterogeneous variance model in Stan offers, at least, two important advantages over frequentist approaches. First, these models can partition the variability multiple times on both the location and the scale side of the model. Thus, inevitably, some of the variance components are going to become small and likely inestimable using a frequentist framework (Asparouhov & Muthén, 2016). As mentioned previously, a Bayesian approach allows researchers to estimate models with small variance components. Second, Stan uses a different sort of Bayesian sampling procedure (i.e., **Hamiltonian Monte Carlo estimation**, specifically the **No-U-Turn Sampler**; Hoffman & Gelman, 2014; see Appendix C for additional explanation) that is more efficient than Metropolis-Hastings sampling and more flexible than Gibbs sampling (i.e., other commonly used sampling procedures in Bayesian analyses).

In addition to describing this estimation procedure, researchers should provide evidence that the sampler has converged and adequately sampled throughout the posterior distribution (Kruschke et al., 2012). For each example, we provide evidence that the different starting values (i.e., different **chains**) converge to the same distribution and that autocorrelation is not playing too large of a role in the results. Fifth and finally, for each example, we provide detailed interpretations of the posterior distribution. This step-by-step process should serve as a guide for researchers estimating Bayesian MELS and heterogeneous variance models.

Illustrative Examples

In this section, we provide illustrative examples for nested cross-sectional and longitudinal designs. These examples utilize simulated data (see Appendix D for details regarding data generation) to demonstrate the sorts of research questions that can be addressed using MELS and heterogeneous variance models. Given that the purpose of this article is to demonstrate the flexibility of these

methods, the illustrative examples involve relatively complex models that incorporate random slopes. In Appendix E, we provide model alternatives that gradually increase in complexity from an empty linear regression model to a MELS model with random slopes on both sides of the model. These buildup analyses are presented using general notation (i.e., not in the context of the illustrative examples). In the remainder of this section, we briefly describe illustrative research questions and then present the analysis and results as would be typical in a published empirical study. In the supplemental material, we provide the code needed to generate the data that serve as the basis for our illustrative examples as well as the code needed to analyze these data. We also provide interpretations of all parameter estimates (see “Readme.doc” for the intended use of each file). The level of detail provided in the supplemental material would not typically be provided in published articles; this information is meant to aid researchers as they move from output interpretation to writing up the results of their own study.

Nested Cross-Sectional Example: LMX Differentiation

This first example illustrates how the MELS model can be used with nested cross-sectional data to provide new insights into the nature of *leader membership exchange (LMX) differentiation*. It is common for supervisors to develop different quality relationships with the subordinates they supervise (i.e., who are nested within the leader’s group) resulting in a pattern of LMX that can be described by central tendency (e.g., the mean), variation (e.g., the standard deviation), and relative position (i.e., how does an individual’s LMX compare to others with the same leader; Martin, Thomas, Legood, & Dello Russo, 2018). The MELS model allows researchers to predict these three properties in a single model, while appropriately considering model uncertainty and avoiding the use of aggregation and its associated problems. In this illustrative example, we consider the effects of transformational leadership (i.e., a Level 2 predictor) and a leader’s perceived similarity with each subordinate (i.e., a Level 1 predictor) on LMX reported by subordinates.

Previous research has found that perceived similarity is predictive of higher-quality LMX relationships (Liden, Wayne, & Stilwell, 1993). As a Level 1 variable, perceived similarity can have between, within (relative), and contextual effects (Enders & Tofighi, 2007; Kreft, De Leeuw, & Aiken, 1995) on either the location or the scale side of the model.² The within-group level of the model (i.e., Level 1 or subordinate level), which applies to both the location side and the scale side of the model, predicts the relative position piece of the LMX outcomes (see Equation D1 in Appendix D). In this example, γ_{10} denotes the relative similarity effect on the location side of the model, and τ_{10} denotes the same effect on the scale side.

Henderson et al. (2009) proposed that transformational leadership may combine with perceived similarity to predict LMX differentiation. They suggest that more transformational leaders (e.g., Leader A in Figure 4) will be less affected by perceived relative similarity. Specifically, we expect that more transformational leaders will have higher mean LMX ratings and that their LMX ratings will be less variable regardless of perceived similarity. In contrast, we hypothesize that less transformational leaders will have lower mean LMX and that their ratings will be more variable (i.e., see differentiation or distance between the data and line for the subordinates who report to Leader B in Figure 4). For these leaders, relative similarity will be more likely to affect relative LMX ratings in terms of the predicted mean LMX as well as in terms of the variation around that predicted mean. Leaders low on transformational leadership will have higher and more consistent LMX with similar group members. These effects manifest themselves in Figure 4 for Leader B as a positive slope of relative similarity on relative LMX and a shrinking distance between the x symbols and the predicted mean (demarcated by the line) as similarity increases. We examine these assertions in our illustrative analysis.

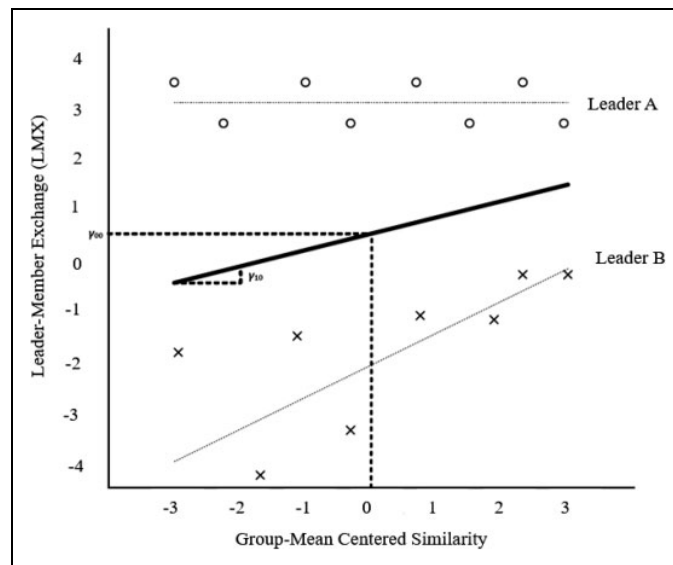


Figure 4. Depiction of hypothetical leaders who differ in transformational leadership. The two lines depicted in this figure represent different leaders (A & B). The location of the subordinates (depicted by o) around Leader A's line (who is high in transformational leadership) indicates that this leader has high and consistent leader-member exchange (LMX) relationships with subordinates regardless of the level of similarity between the leader and the subordinate. Consistency is indicated by the equal distance of the o symbols from the line. The location of the subordinates (depicted by x) around Leader B's line (who is low in transformational leadership) indicates that LMX for this leader depends upon similarity such that the leader has higher LMX with subordinates who are more similar to the leader. Further, the distance of the xs from the line indicates that subordinates who are more similar to the leader are treated more consistently.

Analysis and Results

Prior Justification. To assess the sensitivity of the inferences to the prior, we utilized multiple sets of priors (i.e., two noninformative or vague options and one informed option). The results did not vary greatly. Thus, we present the results using the default uninformed priors. We provide the specific priors used in these analyses below and refer readers to McElreath (2016) and Bürkner (2017) for more information regarding the choice of priors. One advantage of using brms with uninformed priors is that the user does not have to specify the priors. As a default, priors designed not to influence the final results are incorporated into the analysis. In the supplemental material, we illustrate that in less complex models where both the frequentist and Bayesian models are estimated, the results match to two or three decimal places when using similar priors. Thus, these priors are not affecting the final results. In this analysis, **improper flat priors** were used for the fixed effects (i.e., the distribution does not integrate to one, and every possible parameter value is treated as equally likely; Bürkner, 2017). The default prior of a half Student-*t* distribution with 3 degrees of freedom and a location and scale parameter of 0 and 10 respectively was used for the standard deviations of the random effects. The prior for the 4×4 correlation matrix is the LKJcorr (1) (Lewandowski, Kurowicka, & Joe, 2009), which is a flat prior over all valid correlation matrices. The analysis model is the same as the model that was used to generate the data (see Appendix D).

Model Convergence. Before interpreting the results from the MCMC process, we assessed whether the sampler converged and adequately sampled throughout the posterior distribution. If this is not the case, then the results cannot be interpreted. Two potential problems that could prohibit reaching these goals are **start value sensitivity** and high autocorrelation among the sampled parameter

estimates. Like any estimation or sampling procedure, the MCMC sampler used by Stan (i.e., the No-U-Turn Sampler) needs starting values; thus, four chains with different starting values are used (see Appendix B for more information about Bayesian in general and Appendix C for the No-U-Turn Sampler in particular). Moreover, to ensure that potentially poor starting values did not influence the posterior distribution estimates, many of the initial estimates were discarded (i.e., the first half of the 5,000 samples taken from each chain in this study, 2500). The **potential scale reduction factor** (\hat{R}) was used to assess whether the chains converged to the same posterior distribution; \hat{R} was less than 1.05 for all parameters. Thus, the chains are considered to have converged (Asparouhov & Muthén, 2010; see Table 3).

To assess whether autocorrelation is prohibiting adequate sampling from the entire posterior distribution, we investigated two sources of information (i.e., trace plots and the **effective sample size**). The trace plots displaying the sampled estimates for each iteration have a slope of approximately zero; thus, there is no systematic pattern to the autocorrelation (Figure 5). In other words, the plot is indicative of random noise. Additionally, the effective sample size reveals that autocorrelation was not problematic as the ratio of the number of effective samples to the total number of samples taken was greater than .1 (Kruschke, 2015). An additional diagnostic that supports the claim that the sampler performed well is that the posterior predictive plot (Figure 6) also revealed that the model reproduced the data indicating adequate model “fit.”

The final indication that the model is performing as expected is only available because we simulated the data. The credible intervals contained the data-generating parameters for all the parameters except for the location intercept. Thus, the 95% credible intervals contained the data-generating (true) value for 19 of the 20 estimated parameters (i.e., 95%). Also note that the model was able to detect all nonzero fixed effects and the nonzero random effect standard deviations; the ability to detect nonzero effects is important, because having credible intervals that contain all parameters may be an indication that the intervals are too wide. This does not appear to be the case. However, there were some effects that the model did not detect; none of the nonzero correlations between the scale model random effects were significant. Thus, a larger sample size is needed to detect these correlations.

Interpretation of Posterior Distribution. Both the mean and the variance sides of the model contain the similarity-by-transformational leadership cross-level interaction, the associated simple main effects, and the Level 2 similarity effect. The location intercept $\hat{\gamma}_{00} = -.05$, $CI(-.11, -.00)$ is the predicted mean value when all predictors are zero (i.e., at the group mean for perceived similarity, in a group that has an average amount of similarity, and has a leader with an average amount of transformational leadership). The simple main effect of transformational leadership indicates that when similarity is at the group mean, a one-unit increase in transformational leadership results in an expected .37 increase in the LMX mean, $\hat{\gamma}_{02} = .37$, $CI(.32, .42)$. The Level 1 similarity effect on the location side of the model is estimated to be $\hat{\gamma}_{10} = .57$, $CI(.51, .63)$ indicating that subordinates with a leader who has an average amount of transformational leadership are predicted to have a .57 unit mean increase in LMX for each unit increase in similarity. This effect is weakened by the interaction term $\hat{\gamma}_{11} = -.27$, $CI(-.33, -.21)$ (see Figure 7). This effect is interpreted as for each additional unit increase in transformational leadership the effect of similarity on the LMX mean becomes smaller by .27.

The scale side of the model is interpreted similarly. The scale intercept, $\hat{\tau}_{00} = -.14$, $CI(-.19, -.10)$, is the predicted log residual standard deviation for an employee who is at the group mean for perceived similarity, is in a group that has an average amount of similarity, and has a leader with an average amount of transformational leadership. In other words, like any intercept, $\hat{\tau}_{00}$ represents the predicted value when all of the predictors are zero. The simple main effect of transformational leadership indicates that when similarity is at the group mean, the effect of a one-unit increase in transformational leadership results in a .07 decrease in LMX log residual standard deviation,

Table 3. Random and Fixed Effect Output for the Nested Cross-Sectional Leader-Member Exchange Differentiation Example.

Random Effect Standard Deviations and Correlations	brms Output Label	$\hat{\mu}$	$\hat{\sigma}$	$Q_{.025}$	True Value	$Q_{.975}$	Cover	Effective Sample Size
$\hat{\sigma}_{u_0}$	sd(Intercept)	0.25	0.03	0.18	0.23	0.32	1	3567
$\hat{\sigma}_{u_1}$	sd(similarL1_groupmc)	0.22	0.04	0.14	0.22	0.30	1	2775
$\hat{\sigma}_{v_0}$	sd(sigma_Intercept)	0.09	0.04	0.01	0.10	0.17	1	1075
$\hat{\sigma}_{v_1}$	sd(sigma_similarL1_groupmc)	0.05	0.03	0.00	0.10	0.12	1	3019
$\hat{\sigma}_{u_0, u_1}$	cor(Intercept, similarL1_groupmc)	-0.24	0.19	-0.60	-0.30	0.16	1	2577
$\hat{\sigma}_{u_0, v_0}$	cor(Intercept, sigma_Intercept)	0.02	0.32	-0.62	-0.30	0.65	1	4394
$\hat{\sigma}_{u_1, v_0}$	cor(similarL1_groupmc, sigma_Intercept)	0.18	0.36	-0.60	0.30	0.81	1	2918
$\hat{\sigma}_{u_0, v_1}$	cor(Intercept, sigma_similarL1_groupmc)	0.25	0.41	-0.64	0.00	0.88	1	6928
$\hat{\sigma}_{u_1, v_1}$	cor(similarL1_groupmc, sigma_similarL1_groupmc)	-0.06	0.42	-0.81	-0.30	0.76	1	8784
$\hat{\sigma}_{v_0, v_1}$	cor(sigma_Intercept, sigma_similarL1_groupmc)	0.05	0.43	-0.78	-0.30	0.82	1	6853
Fixed or Population-Level Effects	brms Output Label	$\hat{\mu}$	$\hat{\sigma}$	$Q_{.025}$	True Value	$Q_{.975}$	Cover	Effective Sample Size
$\hat{\gamma}_{00}$ Location Intercept	Intercept	-0.05	0.03	-0.11	0.00	-0.00	0	8890
$\hat{\tau}_{00}$ Scale Intercept	sigma_Intercept	-0.14	0.02	-0.19	-0.14	-0.10	1	3980
$\hat{\gamma}_{10}$ Similarity location	similar_groupmc	0.57	0.03	0.51	0.56	0.63	1	9411
$\hat{\gamma}_{01}$ Similarity Level-2 location	similarL2	0.64	0.06	0.53	0.56	0.76	1	9743
$\hat{\gamma}_{02}$ Transformational location	transform_grandmc	0.37	0.03	0.32	0.37	0.42	1	8897
$\hat{\gamma}_{11}$ Location Interaction	similarL1_groupmc: transform_grandmc	-0.27	0.03	-0.33	-0.28	-0.21	1	9926
$\hat{\tau}_{10}$ Similarity scale	sigma_similarL1_groupmc	-0.17	0.03	-0.22	-0.15	-0.12	1	6364
$\hat{\tau}_{01}$ Similarity Level-2 scale	sigma_similarL2	-0.16	0.04	-0.25	-0.15	-0.07	1	12878
$\hat{\tau}_{02}$ Transformational scale	sigma_transform_grandmc	-0.07	0.02	-0.10	-0.10	-0.03	1	9830
$\hat{\tau}_{11}$ Scale Interaction	sigma_similarL1_groupmc: transform_grandmc	0.10	0.02	0.06	0.08	0.15	1	10634

Note: The brms Output Label would not typically be provided in this kind of table; however, we provide it here so that readers can easily link the output from the analysis to the information presented in the table. $\hat{\mu}$ is the posterior mean, $\hat{\sigma}$ is the posterior standard deviation, $Q_{.025}$ is the .025 quantile of the posterior distribution or the lower limit of the 95% credible interval, $Q_{.975}$ is the .975 quantile of the posterior distribution or the upper limit of the 95% credible interval. True value refers to the value that was used to generate the data. Effective sample size is an estimate of the amount of independent information in the auto-correlated samples. \hat{R} is the square root of the total variance divided by the within chain variance and is equal to 1.00 for all parameter estimates. Coverage is whether the quantiles contain the true value – 1 indicates that the interval does contain the true value. σ s correspond to standard deviation when they contain only a single subscript. For example, $\hat{\sigma}_{v_1}$ is the standard deviation of the scale side random slope for similarity. σ s with double subscripts are used to indicate correlations. For example, $\hat{\sigma}_{u_1, v_1}$ is the estimated correlation between the location side and the scale side random slopes.

$\hat{\tau}_{02} = -.07$, $CI(-.10, -.03)$. In other words, transformational leaders tend to have more consistent (or less variable) LMX relationships with their subordinates. The similarity effect $\hat{\tau}_{10} = -.17$, $CI(-.22, -.12)$ indicates that when transformational leadership is at its grand mean, the effect of a one-unit increase in similarity is expected to decrease the log residual LMX standard deviation by .17. The cross-level interaction effect weakens the similarity effect by .10, $CI(.06, .15)$. As seen in Figure 8, the slope of similarity decreases (i.e., less negative) for those higher in transformational leadership.

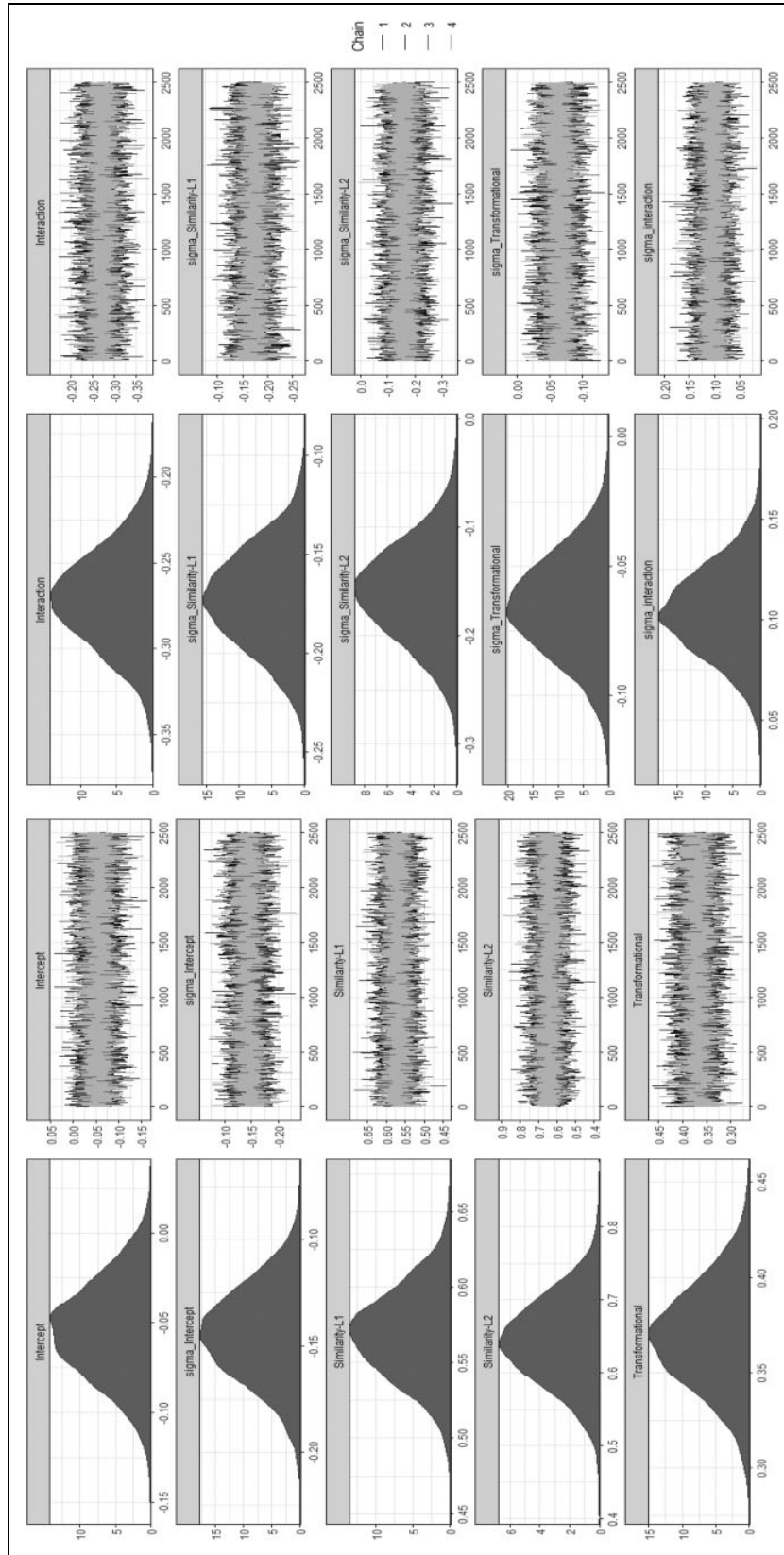


Figure 5. Density and trace plot for the Markov chain Monte Carlo samples for the nested cross-sectional leader-member exchange differentiation example.

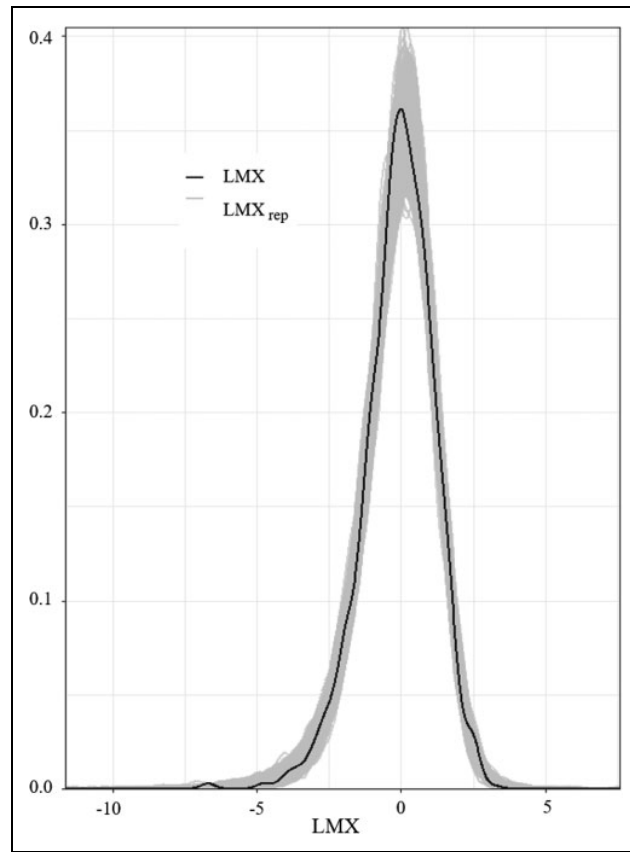


Figure 6. Posterior predictive plot for the nested cross-sectional leader-member exchange differentiation example. The observed data are in black, and the data generated from the model are in gray.

In addition to interpreting the scale side of the model on the log metric, the model can also be interpreted in terms of standard deviations (Lang et al., 2018). These standard deviations are the predicted residual standard deviation (i.e., square root of the residual variance) given specific predictor values. Computing the predicted standard deviation is performed by exponentiating the coefficients; the standard deviations can then be converted into a percentage change that provides a meaningful effect size metric. See Appendix F for a more detailed explanation for this calculation. For example, assuming average similarity (i.e., both the Level 2 and Level 1 predictor values are zero) the predicted standard deviation for a group with a leader that is one unit above average on transformational leadership is calculated as $e^{(-.14 - .07)}$. Thus, their predicted standard deviation value is approximately .81. If the leader had an average level of transformational leadership, then the predicted value would be $e^{-.14}$, which is approximately .87. The percentage change is computed as $(.81 - .87)/.87$, which is approximately $-.07$ or a 7% decrease. Thus, for a one-unit increase in transformational leadership the standard deviation of LMX is reduced by seven percent.

A similar interpretation could be made for the similarity effect for those in a group with a leader that has an average level of transformational leadership and an average group similarity. A one-unit increase in perceived similarity results in a $[(e^{-.14 - .17} - e^{-.14})/e^{-.14}] \times 100 = -15.63\%$ decrease in the predicted standard deviation of LMX. The percentage change is smaller for more transformational leaders. Further, Figures 8 and 9 demonstrate how the slope for similarity as a predictor becomes smaller as transformational leadership increases.

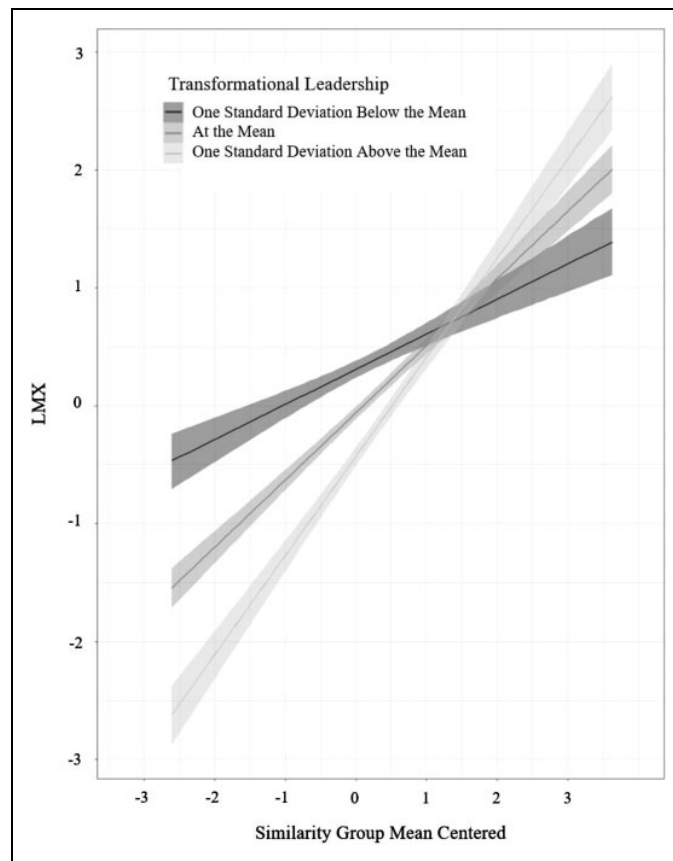


Figure 7. Cross-level interaction plot demonstrating how transformational leadership moderates the effect of similarity on leader-member exchange for the location side of the model (i.e., the predicted mean). The bands reflect the 95% confidence interval.

In summary, LMX relationship quality is affected less by perceived similarity when leaders are higher in transformational leadership. In terms of central tendency, a positive simple main effect of relative similarity (i.e., γ_{10}) indicates that those with relatively more similarity are expected to have higher mean levels of LMX, but that effect decreases as transformational leadership increases. In terms of variation, relatively high similarity results in more consistent LMX relationships. This effect decreases as transformational leadership increases, indicating that transformational leaders treat subordinates consistently regardless of similarity. Groups with similar members have higher mean levels of LMX and less differentiation in LMX ratings.

Longitudinal Example: Between Firm Performance Variability

This second example illustrates how a heterogeneous variance model could be used with longitudinal data to provide new insights into between-firm performance variability differences based on the strategy enacted by different firms. Imitation has traditionally been assumed to reduce between firm performance heterogeneity in an industry. However, Posen and Martignoni (2018) challenged that assumption with a computational model that illustrates the conditions in which imitation can increase between firm performance heterogeneity. They theorize that due to limited information, imitating firms will be unable to duplicate the desired strategy completely and obtain the desired performance. Instead, “copying” firms will have to engage in post-imitation learning where the initial imitation strategy is adjusted. Firms may vary widely

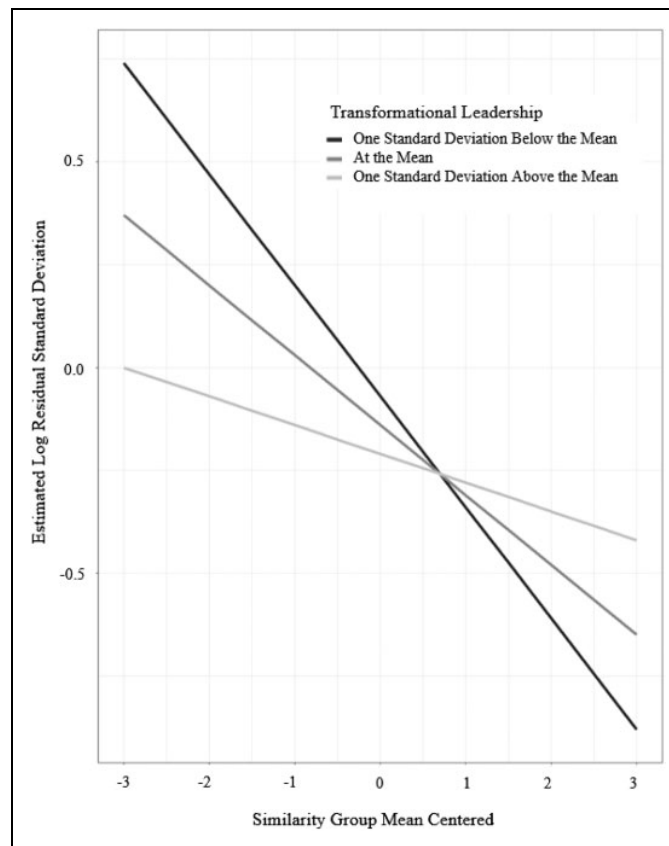


Figure 8. Cross-level interaction plot demonstrating how transformational leadership moderates the effect of similarity on leader-member exchange for the scale side of the model. This plot presents the results for the log residual standard deviation.

in their post-imitation learning ability. Thus, imitation results in both an increased likelihood of achieving the desired performance and an increased likelihood of performing far worse. Thus, there will be increased between-firm performance variability overtime when firms in an industry adopt an imitation strategy.

The data generated in this hypothetical example are meant to emulate monthly performance data across a fiscal year (see Appendix D for details). Thus, time is nested within firm, and each firm is measured on 12 equally spaced occasions. We compare an “imitator” set of firms to a “business as usual” set of firms (i.e., those that continue to perform their preexisting strategies). The outcome is defined generically as firm performance. In terms of heterogeneous variance models, this specific model contains a cross-level interaction on the location side (i.e., the same model as seen in the nested cross-sectional example) and a heterogeneous pattern of Level 2 covariance matrices (i.e., covariance matrices that are systematically varying due to the strategy adopted). The easiest way to estimate this model that allows us to use brms (instead of utilizing Stan directly) is to employ a multiple group approach where each group (i.e., imitators versus business as usual) is allowed to have its own covariance matrix. This approach is similar to some of the approaches described by Kuppens and Yzerbyt (2014). As previously stated, there is a random intercept and linear slope for time as well as their covariance on the location side of the model. In other words, the model allows for differences in terms of initial firm performance (i.e., at the beginning of the fiscal year) and also differences in terms of how much they grow from month to month. The heterogeneous between-group covariance matrices are based on the following rationale.

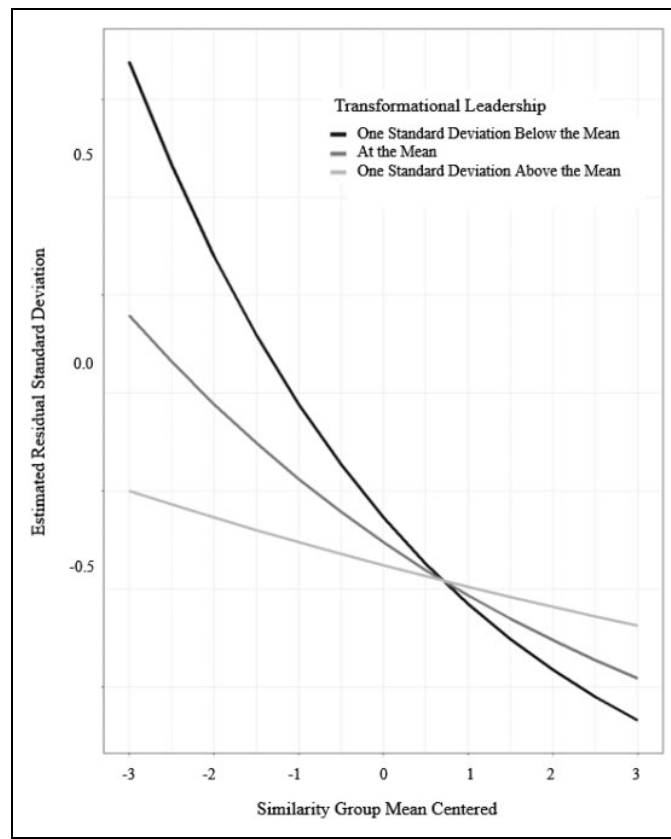


Figure 9. Cross-level interaction plot demonstrating how transformational leadership moderates the effect of similarity on leader-member exchange for the scale side of the model. This plot is rescaled to represent the residual standard deviation.

The decision to imitate the industry leader may be a radical change from the existing practices, and the decision-makers in a firm are likely to be resistant to such radical forms of change. This drastic step would not be taken unless the firm is performing poorly. Based on this rationale, at the beginning of the fiscal year, the imitating firms are hypothesized to have consistent poor performance (i.e., a smaller between-group variability). The location slope variability is hypothesized to be larger in the imitating firms for reasons described previously. Finally, the correlation (i.e., standardized covariance) between the intercept and slope is generated to be larger in the imitation group than the business as usual group. This hypothesis is based on the rationale that very poor initial performance is expected to be associated with a lack of resources or ability to compete in the industry. In other words, no matter what sort of strategy they choose, they are unlikely to succeed.

Analysis and Results

Prior Justification. As in the first illustrative example, the analysis model is the same as the data-generating model, and we use noninformative default priors. The default noninformative priors for the fixed effects are improper (i.e., the distribution does not integrate to one) flat priors (Bürkner, 2017). The standard deviations of the random effects in brms have a default prior of a half student-t distribution with 3 degrees of freedom and a location and scale parameter of 0 and 10 respectively. The prior for each of the 2×2 correlation matrices is the LKJcorr (1) (Lewandowski et al., 2009), which is a flat prior over all valid correlation matrices.

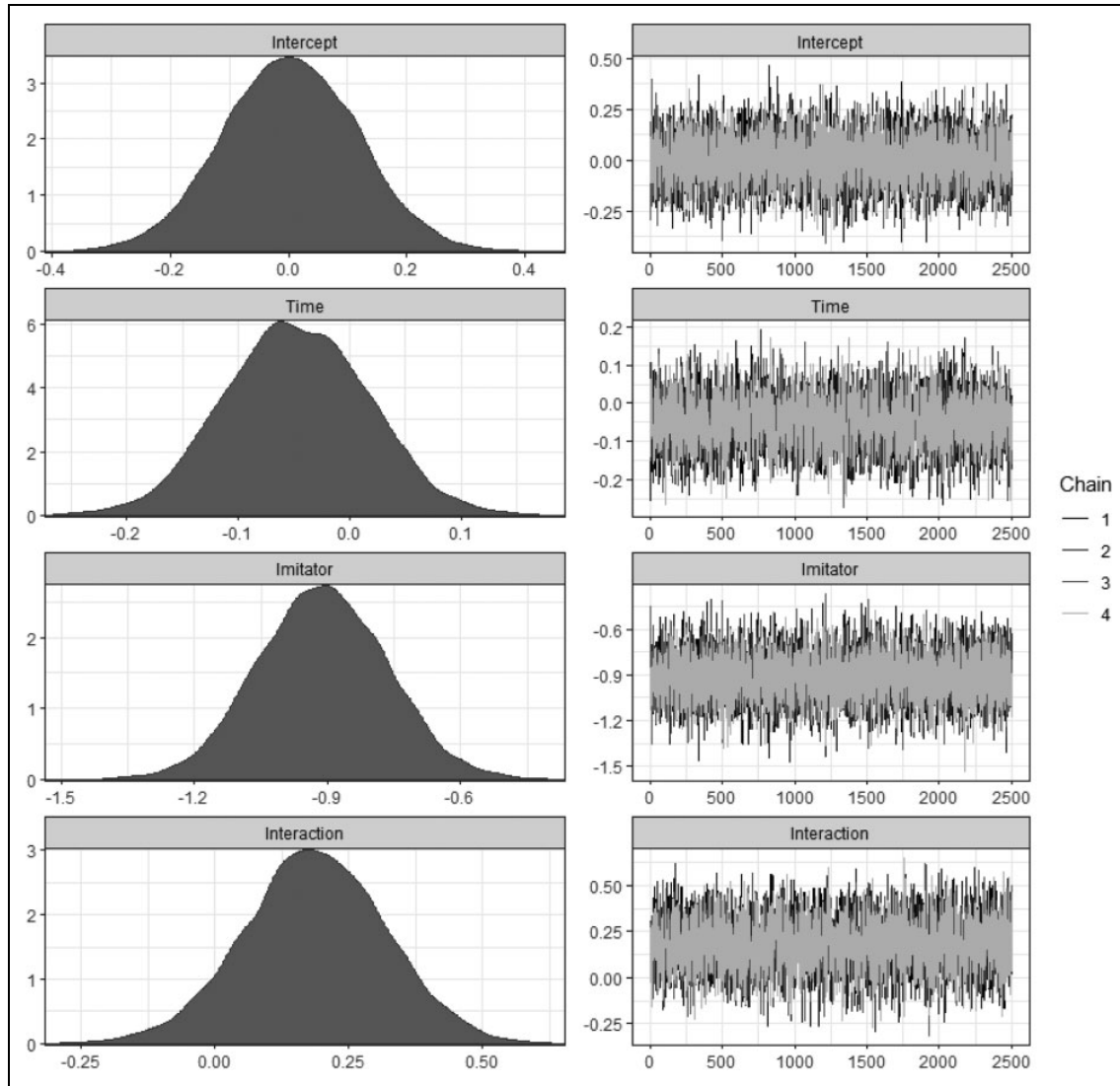


Figure 10. Density and trace plot for the Markov chain Monte Carlo samples for the longitudinal between firm performance variability example.

Model Convergence. The convergence diagnostics of \hat{R} being less than 1.05 for all the parameters estimates suggest that the sampler has converged (see Table 4). The effective sample size ratio computed as the number of effective samples divided by the total number of samples (i.e., $n_{\text{eff}}/10,000$) being larger than .1 or $n_{\text{eff}} > 1,000$ for all the parameters suggests that autocorrelation among the sampled estimates is not causing the sampling procedure to get stuck in any portion of the posterior (see Table 4). The trace plots (Figure 10) indicate that there are no problems with autocorrelation because the slope of those lines is zero and the plots are indicative of random noise. Finally, the posterior predictive plot (Figure 11) also reveals that the model reproduces the data indicating that the model “fits” the data. The final indication that the model is performing as expected is only available because we simulated the data. The credible intervals contained the data-generating parameters for all the parameters.

Interpretation of Posterior Distribution. The interpretation for the location side of the model is the same as a typical multilevel model with a cross-level interaction. The only nonzero effect in this example

Table 4. Fixed and Random Effect Output for Longitudinal Between Firm Performance Variability Example.

				$\hat{\mu}$	$\hat{\sigma}$	$Q_{.025}$	True Value	$Q_{.975}$	coverage	Effective Sample Size
brms Output Label										
Random	$\hat{\sigma}_{u_0}$	Control	sd(Intercept: strat0)	1.46	0.10	1.27	1.41	1.65	1	6251
	$\hat{\sigma}_{u_1}$	Control	sd(time: strat0)	1.02	0.05	0.93	1.00	1.11	1	4378
	$\hat{\sigma}_{u_0}$	Imitator	sd(Intercept: strat1)	.99	0.10	0.81	1.00	1.18	1	4303
	$\hat{\sigma}_{u_1}$	Imitator	sd(time: strat1)	1.77	0.08	1.62	1.73	1.94	1	3629
	$\hat{\sigma}_{u_0, u_1}$	Control	cor(Intercept: strat0, time: strat0)	0.41	0.07	0.26	0.30	0.54	1	1400
	$\hat{\sigma}_{u_0, u_1}$	Imitator	cor(Intercept: strat1, time: strat1)	0.55	0.08	0.39	0.50	0.71	1	1616
	$\sigma_{e_{ij}}$		sigma	1.89	.02	1.85	1.90	1.93	1	11125
Fixed	$\hat{\gamma}_{00}$	Location Intercept	Intercept	0.00	0.11	-0.22	0	0.23	1	8965
	$\hat{\gamma}_{10}$	Location Time	time	-0.05	0.06	-0.17	0	0.08	1	4179
	$\hat{\gamma}_{01}$	Location Imitator	strat	-0.91	0.15	-1.19	-1.00	-0.62	1	9244
	$\hat{\gamma}_{11}$	Location Interaction	time: strat	0.19	0.13	-0.08	0	0.44	1	5423

Note: The brms output label would not typically be provided in this kind of table; however, we provide it here so that readers can easily link the output from the analysis to the information presented in the table. $\hat{\mu}$ is the posterior mean, $\hat{\sigma}$ is the posterior standard deviation, $Q_{.025}$ is the .025 quantile of the posterior distribution or the lower limit of the 95% credible interval, $Q_{.975}$ is the .975 quantile of the posterior distribution or the upper limit of the 95% credible interval. True value refers to the value that was used to generate the data. Effective sample size is an estimate of the amount of independent information in the autocorrelated samples. \hat{R} is the square root of the total variance divided by the within chain variance and is equal to 1.00 for all parameter estimates. Coverage is whether the quantiles contain the true value – 1 indicates that the interval does contain the true value. σ s correspond to standard deviation when they contain only a single subscript. For example, $\hat{\sigma}_{u_1}$ is the standard deviation of the location side random slope for similarity. σ s with double subscripts are used to indicate correlations. For example, $\hat{\sigma}_{u_0, u_1}$ is the estimated correlation between random intercepts and slopes.

was the simple main effect of being an imitator. In line with the first, hypothesis, at the beginning of the study, the imitators are predicted to perform worse than the business as usual condition, $\hat{\gamma}_{01} = -.91, CI (-1.19, -.62)$, see Figure 12. The scale side of the model in this example is incorporated through a multigroup analysis. In this example, the variability differences between the groups are of primary substantive interest. As seen in Table 4, the intercept variability is smaller (32.19% smaller; $[.99 - 1.46]/1.46$) for the imitator group and slope variability is larger (73.53% larger; $[1.77 - 1.02]/1.02$). One-tailed hypothesis tests also reveal that the intercept variability is smaller in the imitator group and that the slope variability is larger (i.e., the credible interval of their difference did not contain zero; see Table 5). Support for the hypothesized correlation difference was not found, because the credible interval contained zero. Thus, these results adhered to our expectations regarding the standard deviations but not for the correlations. At the beginning of the study, the imitators performed consistently worse. Throughout the year some of the firms improved substantially, while others did not (see the relatively large slope variability in Figure 13).

Discussion

Interest in predicting variance-based constructs is increasing among organizational researchers. Lang et al. (2018) recently illustrated the use of a multilevel model to predict changes in consensus (agreement) among members of a group over time. We build upon this first application of multilevel models for predicting heterogeneous residual variances in the organizational sciences by highlighting the flexible nature of MELS and heterogeneous variance models. Researchers can utilize these models in both micro and macro organizational research with different designs (nested cross-sectional or longitudinal) to model different portions of the variability (within- or between-group variability). Below we summarize the advantages of using the most general MELS models.

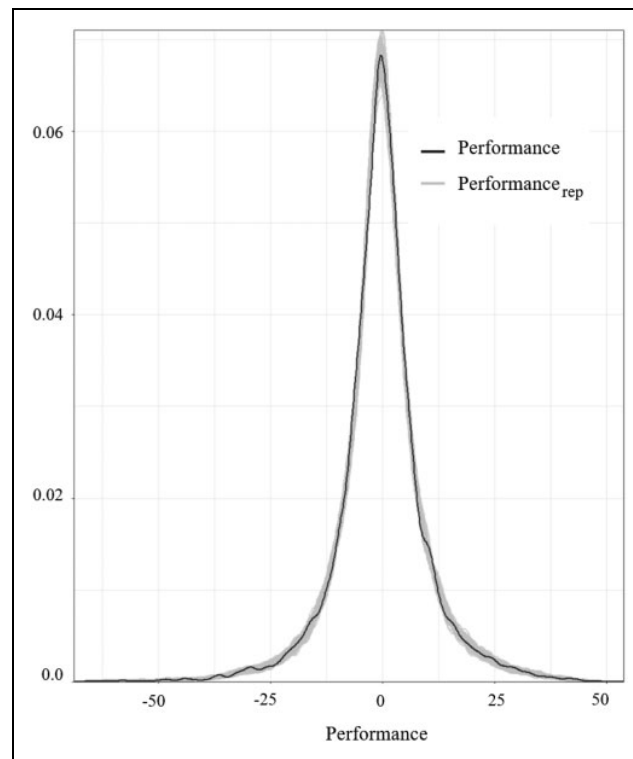


Figure 11. Posterior predictive plot for the longitudinal between firm performance variability example. Note that the observed data are in black and the data generated from the model are in gray.

Advantages of Using MELS

The MELS model offers several statistical advantages over other approaches, including heterogeneous variance models. First, by including random effects (particularly the random intercept on the scale side of the model), it is possible for researchers to test the assumption that all Level 2 units have the same amount of within-group/person variability (i.e., homogeneous variances) and avoid inflated Type I error rates (Leckie et al., 2014; Walters et al., 2018). Moreover, this model allows researchers to avoid the limitations of aggregation (i.e., computing a standard deviation for each group and predicting it in a single level regression) that include the inability to include Level 1 predictors and disregarding uncertainty in the aggregated variability estimates. This uncertainty should be and is accounted for when the data are analyzed with the MELS model.

As mentioned previously, applying MELS models to predict a construct of substantive interest increases the types of research questions that researchers can address. To this point, organizational researchers have predominately focused on hypotheses about the amount (mean) of responses. However, researchers should consider whether there are sound theoretical reasons to propose hypotheses aimed at predicting variability. Our illustrative example using nested cross-sectional LMX data exemplifies the types of questions that can be addressed with multilevel models but are not possible to examine using an aggregation approach. In the example, cross-level interaction research questions with random slopes for both the mean and the residual standard deviation were addressed using data that were generated based upon a substantive example. In the context of LMX and LMX differentiation, perceived similarity was allowed to have random slopes when predicting LMX mean (location) and LMX differentiation (scale); perceived similarity was also allowed to have Level 2 effects on both sides of the model. The effect of perceived similarity was qualified by transformational leadership as a moderator.

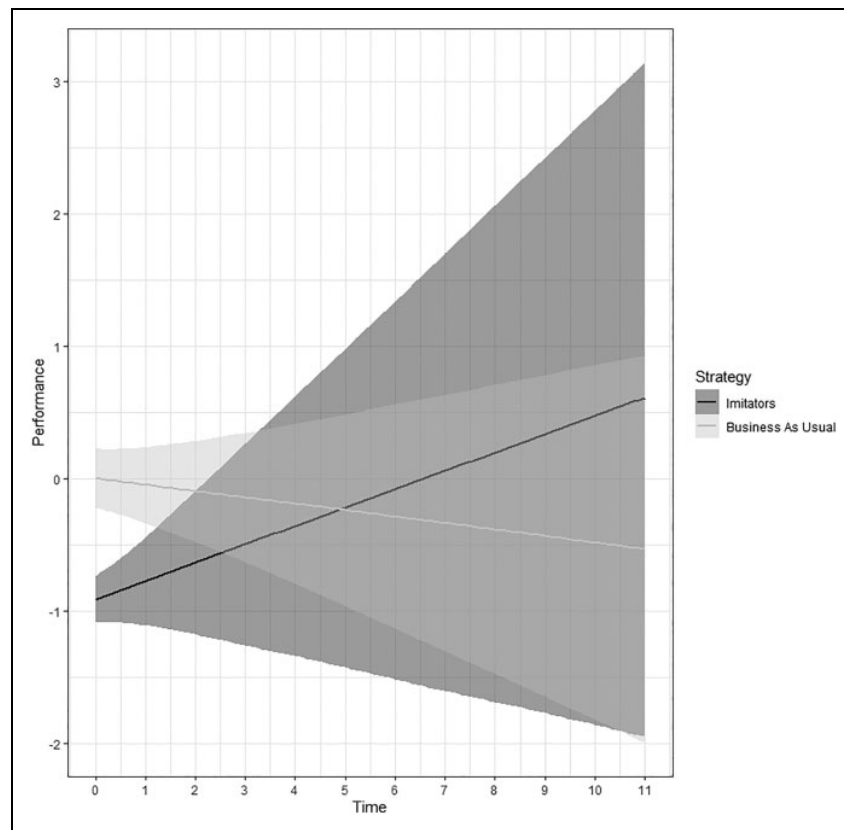


Figure 12. Cross-level interaction plot demonstrating how strategy moderates the change in average performance over time. The shaded regions or bands reflect the 95% confidence interval.

Table 5. Variability Magnitude Hypotheses.

Hypothesis	$\hat{\mu}_{diff}$	$\hat{\sigma}_{diff}$	$Q_{.05}$	$Q_{.95}$
$\hat{\sigma}_{u_0}$ Control – $\hat{\sigma}_{u_0}$ Imitators > 0	0.47	0.15	0.25	∞
$\hat{\sigma}_{u_1}$ Control – $\hat{\sigma}_{u_1}$ Imitators < 0	–0.75	0.09	$-\infty$	–0.60
$\hat{\sigma}_{u_0, u_1}$ Control – $\hat{\sigma}_{u_0, u_1}$ Imitators < 0	–0.15	0.11	$-\infty$	0.03

Note: $\hat{\mu}_{diff}$ is the posterior mean of the differences between the two estimates; $\hat{\sigma}_{diff}$ is the standard deviation of the difference. These are one-sided confidence intervals. Thus, only one side of the credible interval is of interest. When evaluating a less than hypothesis only the upper limit is needed, and when evaluating a greater than hypothesis only the lower limit is needed. Note that the credible intervals do not include zero for the variances but do for correlation comparison.

Researchers may also examine questions that consider the amount (mean) and the variability in combination. For example, researchers studying the results of an intervention may find that a particular treatment not only increases the mean of a construct (e.g., employee wellbeing) but also decreases the variability of ratings over time. As another example, researchers might propose that the intervention treatment group would start with a mix (high variability) in wellbeing scores, but over time, all individuals in the treatment group would both improve and become more homogenous. In contrast, a less advantageous treatment might increase the mean, but may also increase the variability. In this case, although on average wellbeing may be improving, the results also suggest that the treatment is creating a disparity between people. In other words, the treatment works for some but not for others. This finding would be similar to the results of the second illustrative

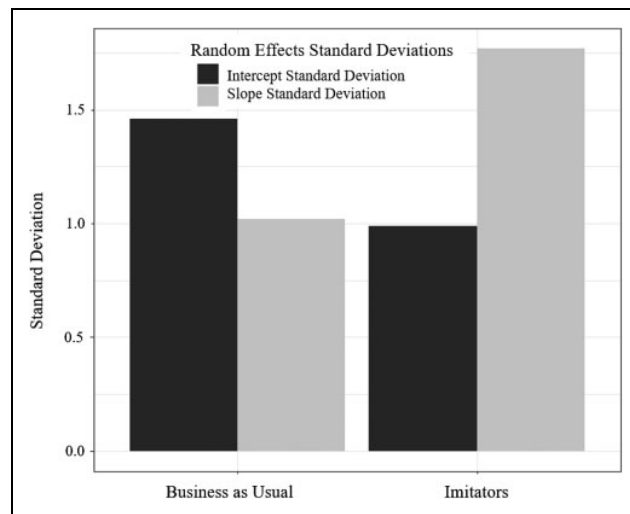


Figure 13. Bar chart displaying random intercept and slope standard deviations by strategy.

example examining firm performance. The ability to predict the mean and variability in the same model makes it possible to test these sorts of combination hypotheses.

Practical Recommendations

Bayesian estimation is appropriate and likely necessary when estimating complex MELS models because of the large number of random effects. The Bayesian approach, in conjunction with brms (Bürkner, 2017) and Stan (Carpenter et al., 2017), utilized in this article allows for the estimation of random slopes as well as cross-level interactions in both sides of the model. In terms of software comparison, brms and Stan is the most widely available alternative. The only other software that has been utilized to estimate this model is Stat-JR which is also a Bayesian program that is distributed with MLWin (Browne, 2017). MixRegLs (Hedeker & Nordgren, 2013) or MixWild does not allow for random slopes on the scale side, and as of version 8.0, Mplus (Muthén & Muthén, 1998-2017) can specify heterogeneous variances, but it is unclear how a random slope could be incorporated on the scale side. Perhaps, the dynamic structural equation modeling framework (Asparouhov, Hamaker, & Muthén, 2018) that incorporates time series analysis into a structural equation modeling framework could provide an alternative.

Bayesian analyses offer several statistical advantages; however, these approaches are also unfamiliar to many researchers. Providing priors is an extra step in conducting an analysis, and researchers must ensure that their choices regarding priors do not influence the final result inappropriately. While brms helps reduce the burden on researchers by supplying default priors that seek not to influence the final results, this is still an issue that researchers should consider. Convergence issues can occur in Bayesian analysis, and when they do, researchers must resolve these issues by investigating various convergence diagnostics. Particularly, researchers may face challenges when estimating random slopes on both sides of the model. We provide code to perform the analyses but do not provide a lengthy discussion of the process of addressing convergence issues. We encourage researchers to consult accessible resources, including Kruschke (2015) and McElreath (2016), as this process is very data/model dependent.

To demonstrate the flexibility provided by utilizing a Bayesian approach to testing variability-related hypotheses, we present a complex MELS model (i.e., with random slopes on the location and scale sides). However, in some cases, less complex models may be more aligned with the questions

of substantive interest. We recommend readers use Appendix E as a tool to consider the different types of models they may estimate and to pick the model that best fits their theory and research questions.

One very important takeaway from our review of the existing literature is that precise language is necessary when interpreting complex multilevel models. We found instances where researchers were conducting traditional multilevel models (e.g., growth models) and interpreting the results as if they were predicting heterogeneous variances. Language such as “we predicted variability” may lead readers to think that the study is predicting heterogeneous residual variances when the study is actually using a traditional multilevel model to explain a portion of a single variance estimate that the model provides for everyone in the study (e.g., a Level 2 predictor explains some random intercept variance; see Figure 3). Thus, we recommend that researchers use more precise language when describing their analysis. For example, researchers applying these models should clarify that they used a MELS or heterogeneous variance model to predict heterogeneous residual variances. This language should make it clear that the outcome of interest is a variance-based construct and that these analyses are distinct from explaining variability as is done in a typical regression analysis.

Researchers may have questions regarding centering decisions in MELS and heterogeneous variance models. To our knowledge, researchers have not examined this issue with these specific models. However, given that the scale-side of the model is a log-linear model, researchers may interpret the τ s as they would the β s on the scale side. Enders and Tofighi (2007) advise that centering decisions should be based on the research question of interest. Researchers who are interested in Level 1 predictors could include the group mean centered version of the Level 1 predictor without including the Level 2 variable. However, if this approach is used, it is important to remember that all Level 2 variability is being ignored. When researchers are interested in both Level 1 and Level 2 versions of the same predictor, then either group or grand mean centering can be used to provide the Level 1, the Level 2, and the contextual effect. Either approach will provide two of these three effects, and the third effect can be calculated from the other two (Enders & Tofighi, 2007). Further, for researchers interested in a Level 2 predictor only, centering decisions are not overly complicated, and advice concerning single-level regression analyses applies.

Researchers may be interested in providing effect size calculations for models with random slopes on the scale side. Pseudo- R^2 values could be calculated for each of the variance components in this model, and they have been shown to work well with models that include random intercepts on both the location side and the scale side (Walters et al., 2018). However, these analyses have yet to be extended to the random slope case. Until this research is conducted, researchers may consider using the intuitive effect size estimate obtained by calculating the proportional increase/decrease in the standard deviation (Lang et al., 2018).

Researchers should be aware that a mediation version of the MELS model has also not yet been developed and evaluated. Many of the hypotheses discussed in this article could easily be expanded to include mediated relationships. For example, the LMX example could be extended to view LMX differentiation as a mediator and add group performance as the distal outcome (Henderson et al., 2009). To address this hypothesis without the shortcomings of aggregation, a mediation version of the model would need to be developed.

Power analyses have also not yet been conducted for the full MELS model presented in this article (i.e., random slopes on both sides of the model). However, recent research examining less complicated models has begun to provide some insight into power issues with these analyses. For example, Walters et al. (2018) found that the power to detect the scale-model random intercept variance and the effect of an individual-level predictor of residual variability increased with the number of individuals and occasions of measurement (and provided power curves that may help researchers planning to conduct a similar analysis).

Research on longitudinal modeling has demonstrated the benefits of model building that occurs in a step-up fashion where the models start simple and become increasingly complex. For example, polynomials of increasing order are often fit to get the best model for improvement over time before adding any additional predictors to the model. Within a given side of the model, it makes sense to model the effect of time correctly before adding in any additional predictors. However, whether researchers should model the location (mean) or scale (variance) side of the model first remains an open question or future research. As methodological work on these models proceeds, researchers should ground their modeling building in theoretical arguments and also test multiple competing models. These models can be compared using the Watanabe-Akaike information criterion (WAIC; Gelman, Hwang, & Vehtari, 2014) which is an extension to the commonly applied Akaike information criterion (AIC; Akaike, 1998) that does not assume that posterior distribution is approximately multivariate normal. We provide examples illustrating how to conduct model comparisons in the supplemental material.

Conclusion

Variability is no longer viewed as something to be averaged over, ignored, or tricked into fitting the assumptions of a typical general linear model. It is substantively interesting to researchers working in many different areas. Regression models produce normal distributions that are conditional on the values of the predictors and their associated regression coefficients. Like any normal distribution, these distributions are characterized by both the mean and the variance. Thus, every effort that is used to model the mean correctly should be used to model the variability correctly as well. By expanding the multilevel repertoire to include MELS and heterogeneous variance models, researchers can appropriately analyze mean and variability-related hypotheses at multiple levels simultaneously.

Appendix A

Glossary

This glossary provides definitions for key terms presented in the main text. Additional terms are defined in the supplemental material (e.g., terms presented in the analysis output).

Substantive Terms

Variance-based constructs: Constructs that are defined by their variability (e.g., standard deviation, variance, range) instead of their mean. For example, pay dispersion within a company is defined by the variability in pay among members, not by the (average) amount of pay.

Multilevel Terms

Fixed effects: Parameters that do not vary across groups/persons. There is a single estimate for the entire population. The brms software labels these **population effects**. The reason for this distinction is discussed in Gelman and Hill (2006). In this article, we use the term *fixed effects*. Fixed effects do not have a group subscript on the coefficients (e.g., γ_{10})

Random effects: Parameters that are allowed to vary across groups/persons. Each group/person receives their own estimate for these parameters. The focus is typically on the mean/variance of these effects. These are labeled group effects by brms. We use the term *random effects* throughout the article. A coefficient that is allowed to vary randomly will have a group subscript (e.g., β_{0j}).

Heterogeneous variance models: This model includes an equation to predict the mean (like any typical general linear model, including t tests, analysis of variance, regression) and also an equation to predict or model heterogeneous variances. Equation 3 in Figure 1 presents the equation for a heterogeneous variance model; this model does not contain random effects on the scale side of the model (i.e., no σ^2). The model results in heterogeneous variance that is systematically varying (i.e., the only reason the variances are heterogeneous are due to our predictors of interest). This model is making the assumption that we know exactly why the variances are heterogeneous (i.e., W_j). The longitudinal example presented in this article is a heterogeneous variance model (predicting Level 2 variance) because it does not have random effects on the scale side of the model.

Mixed-effects location-scale models (MELS): Similar to the heterogeneous variance model, MELS includes an equation to predict the mean and also an equation to predict or model heterogeneous variances. Specifically, this multilevel model includes a model for the mean (location) as well as the (log) residual standard deviation/variance (scale). The distinguishing characteristic of this model is that it contains random effects on both sides of the model (see Equation 4 in Figure 1). This model relaxes the assumption that researchers know exactly why some groups have more within-group variability than other groups by including random effects. The nested cross-sectional example presented in this article is an example of a MELS model.

Bayesian Terms

Prior distributions: Represent the information available about the parameter of interest before the study is conducted. Prior distributions may either be informative (i.e., represent the researchers' beliefs regarding plausible parameter values) or noninformative (i.e., attempt to have a prior that does not affect the results).

Improper flat priors: Improper refers to the fact that the prior distribution does not integrate to a finite value (e.g., 1). If the prior integrates to a finite value other than 1 (e.g., q), then it can be multiplied by its reciprocal ($1/q$) to achieve a proper prior that integrates to 1. Flat indicates that the prior assumes that all values are equally likely that would appear on a graph as a horizontal or flat line. In Stan, parameters that are not assigned a prior are assumed to have an improper flat prior. Please note that in some other Bayesian programs (e.g., JAGS/BUGS) the user must specify a prior distribution for all parameters.

Likelihood: Represents the contribution of the data in Bayesian inference. It is the density (e.g., a normal distribution) for the observed sample for a given parameter value. The likelihood is viewed as a function of the parameter for fixed data and is often denoted as $L(\theta|x) = f(x|\theta)$ (Casella & Berger, 2002) where θ is the parameter vector, x is the observed data, and $f(x|\theta)$ is the joint probability density or probability mass function given θ . This combination, which is obtained by multiplying the prior by the likelihood, is known as the posterior distribution.

Posterior distribution: Posterior distributions reflect the updated plausibility of the parameter values after the likelihood has been combined with the observed data.

Hamiltonian Monte Carlo estimation: Uses a physics simulation to propose new parameter values in the Markov chain Monte Carlo (MCMC) process. The simulation utilizes four pieces of information to conduct this simulation: (a) the numerator of Bayes's theorem, (b) the gradient, (c) the number of steps to be taken, and (d) the length of those steps. The role of each of these pieces of information is explained in Appendix C. For more information, see the following texts: Kruschke (2015), McElreath (2016), and Neal (2010).

No-U-Turn Sampler: The No-U-Turn Sampler is the algorithm used by Stan to ensure that the sampling procedure does not take a U-turn and produce new estimates that are very similar to those produced at the prior iteration. This algorithm is described at length in Appendix C.

Effective sample size: The samples that are generated by the MCMC process are correlated to some degree. Thus, the amount of information in the samples is less than there would be if the sampled estimates were uncorrelated. This results in increased estimation uncertainty. The Stan reference manual describes effective sample size as the number of independent samples with equivalent estimation power as the autocorrelated samples obtained from the MCMC process. The more correlated the samples are the smaller the effective sample size. See https://mc-stan.org/docs/2_18/reference-manual/effective-sample-size-section.html for additional information.

Start value sensitivity: We use this term to indicate that the results obtained from the Bayesian estimation are sensitive to the initial parameter estimates that were provided to the estimation procedure. In other words, where the estimation procedure starts is affecting where it ends up. To assess whether this is the case, multiple independent chains with different starting values can be used to assess whether they arrive at the same posterior distribution.

Chains: The number of chains indicates the number of independent starting points (draws) in the sampling procedure. Within a chain the sample parameter estimates are dependent on one another (autocorrelated); however, between chains the estimates are independent. Multiple chains are used to assess convergence (i.e., that multiple starting points have converged to the same posterior distribution).

Potential scale reduction factor: Used to assess the convergence of a set of simulated chains. It is calculated as the square root of the marginal posterior variance divided by the within-chain variance. Values less than 1.05 are considered satisfactory (Asparouhov & Muthén, 2010).

Appendix B

Introduction to Bayesian Statistics

Like any branch of statistics, Bayesian methods seek to use information from a sample to learn about plausible values for an unknown parameter of interest. Informally, Bayesian statistics provide a mathematical way of learning from experience where experience comes in the form of data (e.g., Kruschke, Aguinis, & Joo, 2012; McElreath, 2016). Learning is reflected by the reallocation of credibility from what was thought originally, referred to as the prior, to the new set of beliefs regarding credible values of the parameter, the posterior. This appendix provides a brief introduction to how Bayesian methods are used to estimate the plausibility of parameter values given data. Several years ago, Kruschke and colleagues (2012) provided an introduction to Bayesian methods and highlighted several general advantages over frequentist approaches. For example, while null hypothesis significance testing tells researchers the “probability of obtaining the data in hand, or a more extreme unobserved data, if the null hypothesis were true . . . [Bayesian approaches tell researchers] the credibility of candidate parameter values given the data that we actually observed” (p. 723). In other words, they provide the information researchers seek. Further, Bayesian statistics provide an entire distribution of parameter estimates as opposed to the single point estimate and confidence interval obtained from frequentist analyses. Importantly, for the models we present, Bayesian approaches enable researchers to successfully estimate models with many random effects (e.g., with random slopes on both the location and the scale side of the model). We refer readers to the article published in this journal by Kruschke and colleagues as well as several introductory texts for Bayesian methods (e.g., Kruschke, 2015; Lynch, 2007; McElreath, 2016) to gain a more complete introduction to these methods.

Bayes's Theorem

Bayes's theorem is the tool that allows researchers to make inferences about the unknown parameters given the data. This section explains how fundamental assumptions of Bayesian approaches lead to the formulation of Bayes's theorem. Both Bayesian and frequentist approaches assume that parameters have a true (fixed) value (Greenland, 2006). Kaplan (2014) describes the fundamental difference between these approaches as assuming that the unknown parameter of interest has a fixed value that researchers aim to estimate (frequentist tradition) versus assuming this parameter is a random variable with a probability distribution that reflects researchers' uncertainty about the parameter's true value (Bayesian tradition). This difference is what allows Bayes's theorem to specify the relationship between the prior distribution, the likelihood, and the posterior distribution.

$$P(\text{parameters}|\text{data}) = \frac{P(\text{data}|\text{parameters})P(\text{parameters})}{P(\text{data})} = \frac{\text{likelihood} \times \text{prior}}{\text{data}} \\ \propto \text{likelihood} \times \text{prior}$$

Prior distributions are specified before data are collected and may be either informative (i.e., represent the researchers' beliefs regarding plausible parameter values) or noninformative (i.e., attempt to have a prior that does not affect the results). It is possible that the choice of noninformative prior could influence model estimates as well as model selection; thus, a sensitivity analysis should be performed to determine the influence of the model priors (Depaoli & van de Schoot, 2015). The **likelihood** is the density (e.g., a normal distribution) for the observed sample for a given parameter value. The likelihood is viewed as a function of the parameter for fixed data and is often denoted as $L(\theta|x) = f(x|\theta)$ (Casella & Berger, 2002) where θ is the parameter vector, x is the observed data, and $f(x|\theta)$ is the joint probability density or probability mass function given θ . This combination, which is obtained by multiplying the prior by the likelihood, is known as the **posterior distribution**. In other words, combining the prior with the likelihood is how Bayesian statistics allow researchers to learn from experience. The posterior distribution is what researchers ultimately use to make inferences and contains the relative credibility of the parameters given the data.

The research containing Bayes's theorem was published in 1763 (Bayes, 1763); however, the method was not regularly applied until the 20th century or even the 21st for some fields (Kruschke et al., 2012). Use of this method was delayed because calculating the denominator of Bayes's rule involves a high-dimensional integral that can be very difficult to compute. Computer-intensive methods (e.g., Markov chain Monte Carlo; MCMC) now allow researchers to approximate the posterior distribution by drawing representative samples from it without needing to calculate the denominator in Bayes's theorem. This is reflected by the final line in the equation above that does not present the denominator. The variant of MCMC that is used to estimate the models presented in the article is referred to as the No-U-Turn Sampler (Hoffman & Gelman, 2014). It is described in detail in the following appendix.

Appendix C

Markov Chain Monte Carlo Appendix

This appendix describes the specific sampling method utilized to approximate the posterior distribution for the Bayesian analyses presented in this article. In the supplemental material, we provide code to utilize Stan to estimate these models. Stan employs a Hamiltonian Monte Carlo (HMC) process, which runs a physics simulation where the current estimate or location of the parameters is viewed as a particle on a frictionless surface (i.e., the negative log posterior). This HMC process

results in a sampling process that is more computationally demanding to obtain a single sampled parameter estimate but is much more efficient, thus requiring fewer samples. Overall, this sampling process takes less time and performs better when models contain more estimated parameters than other sampling procedures (e.g., Metropolis-Hastings or Gibbs sampling procedures commonly utilized in Bayesian analyses conducted in OpenBUGS, JAGS, and Mplus; McElreath, 2016). A limitation of HMC sampling is that this process is designed for continuous parameters, but that is not a problem for the multilevel models discussed in this article.

Four pieces of information are needed to select each sampled value in the HMC process: (a) the numerator of Bayes's theorem, (b) the gradient (i.e., a vector of partial derivatives where the numerator of Bayes's theorem is differentiated with respect to each parameter; in other words the slope in each direction given the parameter estimates' current location), (c) the number of steps to be taken, and (d) the length of those steps. The numerator of Bayes's theorem provides researchers with information regarding the height given a set of parameter values. The gradient allows researchers to know the slope in all directions given a current position (i.e., set of parameter estimates). The particle or set of parameter estimates is given a random push that is random in terms of both force and direction. The curvature of the surface (i.e., assessed by the gradient) determines how far the particle moves.

Because the surface is frictionless, researchers must decide when to stop the physics simulation. This is where the number of (c) and length of the steps (d) come into play. These two settings are determined automatically during the "warm-up" phase by Stan. Smaller step sizes result in better approximations to the continuous process but can make the simulation take a long time to move from its current position. If the number of steps and step size combination is inappropriate, the sampling process may "turn around" and return to its previous location. This is what is referred to as the U-turn problem that results in very high correlation between the parameter estimates. Fortunately, the number of steps is tuned automatically by Stan using a process that determines when the parameter estimates are beginning to turn around to return to their previous location. The algorithm utilized by Stan is called the No-U-Turn Sampler (Hoffman & Gelman, 2014). See the following texts for more information regarding these samplers (ordered in terms of increasing technical detail): Kruschke (2015), McElreath (2016), and Neal (2011).

Appendix D

Data Generation

The R code used to generate the data for the examples is provided in the online supplemental material. We also provide the generated data. Please follow this link to access these materials: https://osf.io/4ves7/?view_only=3a4c5e5f9df14a268cfb5d9392bb539e.

In each example, the analysis model is the same as the data-generating model. Thus, the credible intervals (i.e., the Bayesian term for confidence interval) should (and do) contain the parameter values.

Nested Cross-Sectional Example

To the extent possible, the coefficients obtained to generate these data are from meta-analytic estimates (Dulebohn, Bommer, Liden, Brouer, & Ferris, 2012). The correlations from the meta-analyses were converted to pseudo-R² values for each respective variance component. Because the MELS model has not been used to study leader-member exchange (LMX) relationships, there was no prior information for some pieces of the model. When prior information was not available the coefficients were chosen to reflect effect sizes typically found in organizational research. The lack of

prior information from the extant literature to inform the generation of portions of this model is not detrimental because these data and analyses are intended to be illustrative.

Data Generation

Two hundred fifty teams with sizes ranging from three to ten were generated ($M = 6.94$, $SD = 1.46$). The equation form of this data-generating model is built on the common notation made popular by Raudenbush and Bryk (2002) and similar to that presented in Leckie, French, Charlton, and Browne (2014). The data-generating location model is shown below.

$$\begin{aligned} LMX_{ij} &= \beta_{0j} + \beta_{1j}(\text{Similarity}_{ij} - \overline{\text{Similarity}}_j) + e_{ij} \\ \beta_{0j} &= \gamma_{00} + \gamma_{01}\overline{\text{Similarity}}_j + \gamma_{02}(\overline{\text{Transformational}}_j - \overline{\overline{\text{Transformational}}}) + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11}(\overline{\text{Transformational}}_j - \overline{\overline{\text{Transformational}}}) + u_{1j} \end{aligned} \quad (D1)$$

The Level 1 predictor was perceived similarity (group-mean-centered) and the Level 2 predictors were the Level 2 similarity effect and transformational leadership which was grand-mean centered. Grand-mean centering was performed to ensure the interpretability of the similarity simple-main effects. The predictor variables were generated to be standard normal and assumed to be uncorrelated with one another; the regression coefficients were calculated to account for the proportion of variability that corresponds to the meta-analytic correlation estimates (i.e., $.73^2 = .53$ and $.5^2 = .25$ for the relationship between LMX and transformational leadership and perceived similarity, respectively; Dulebohn et al., 2012). The Level 1 and Level 2 relationship between LMX and similarity was held constant. Additional details are provided in the data generation code; Table 3 contains the data-generating values in the column labeled “True Value.”

The residual values e_{ij} in the location model are traditionally assumed to be independent and identically distributed with a mean of 0 and some constant variance or standard deviation σ_e . The MELS model relaxes the assumption of constant residual standard deviation by including fixed and random effects to predict its value. The data-generating scale model for the residual standard deviation is shown below, which uses the (natural) log link to ensure that predicted residual standard deviation remains positive as it should.

$$\begin{aligned} \log(\sigma_{e_{ij}}) &= \alpha_{0j} + \alpha_{1j}(\text{Similarity}_{ij} - \overline{\text{Similarity}}_j) \\ \alpha_{0j} &= \tau_{00} + \tau_{01}\overline{\text{Similarity}}_j + \tau_{02}(\overline{\text{Transformational}}_j - \overline{\overline{\text{Transformational}}}) + v_{0j} \\ \alpha_{1j} &= \tau_{10} + \tau_{11}(\overline{\text{Transformational}}_j - \overline{\overline{\text{Transformational}}}) + v_{1j} \end{aligned} \quad (D2)$$

Both the location and the scale model have random intercepts and slopes that can be allowed to relate to one another. The random effects are assumed to be distributed multivariate normal with a zero mean vector and a covariance matrix. Thus, the covariance matrix containing these relationships is 4×4 in this example (see Equation 5 in the article). These estimates are interpreted like any other covariance/correlation matrix. For example, a negative correlation between the location and scale random slopes would indicate that leaders who have higher quality relationships with more similar subordinates are likely to treat those subordinates consistently.

The average residual variance was chosen to ensure that the unconditional (empty model with no predictors) intraclass correlation (ICC) was within the range typically found in the organizational sciences (i.e., between 0.15 and 0.30; Mathieu et al., 2012); the ICC for the generated data was estimated to be 20.36% (i.e., .2036). Figure D1 contains the spaghetti plot of the generated data where each group is represented by their own line and each data point is a subordinate's score.

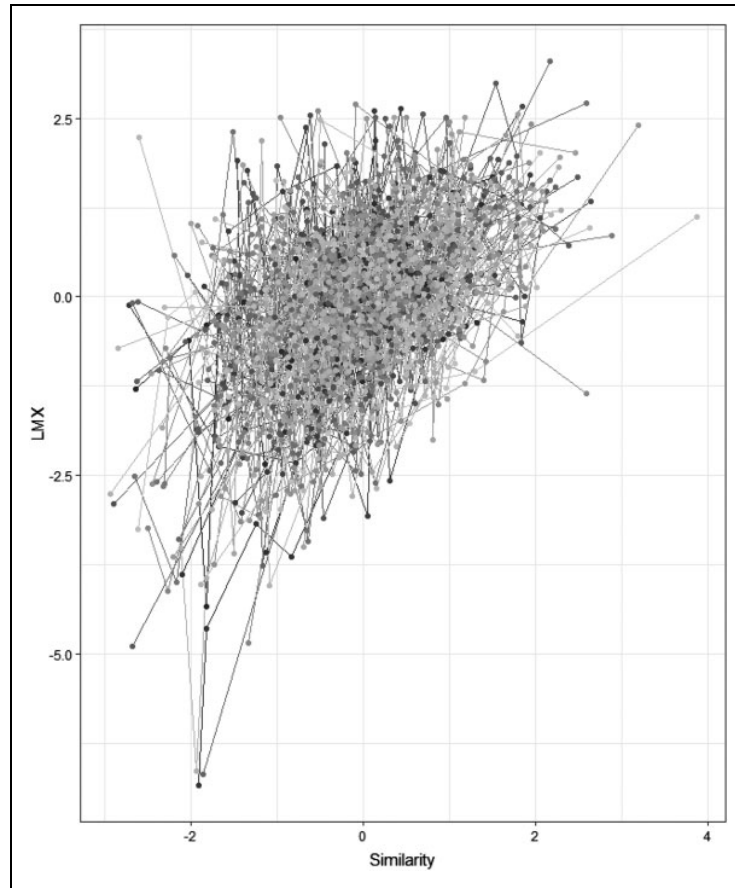


Figure D1. Spaghetti plot for the nested cross-sectional LMX differentiation example.

Longitudinal Example

A cross-level interaction model was used for data generation on the mean side of the model (see equation below). The inspiration for this longitudinal example was drawn from a computational model that was not tied to any particular industry (Posen & Martignoni, 2018). Industry-specific information could play a large role in the values that the coefficients take. Thus, to keep the inferences general, the coefficients chosen are meant only to reflect the hypothetical substantive description provided earlier and not to be industry-specific. These data and analyses are intended to illustrate the sorts of models and inferences that can be made using a heterogeneous variance model. The heterogeneity of variance in this model takes the form of heterogeneous Level 2 covariance matrices. In other words, the entire covariance matrix is allowed to vary across the imitator and control conditions. The specific values of the variances are provided in Table 4. Because the focus is on the Level 2 variances, 500 firms were generated in this example. Each firm has monthly performance data for a year; thus, the 12 time points are viewed as nested within firm. The spaghetti plot of the data plotting firm performance over time is presented in Figure D2.

$$\begin{aligned}
 \text{Performance}_{ij} &= \beta_{0j} + \beta_{1j} \text{Time}_{ij} + e_{ij} \\
 \beta_{0j} &= \gamma_{00} + \gamma_{01} \text{Strategy} + u_{0j} \\
 \beta_{1j} &= \gamma_{10} + \gamma_{11} \text{Strategy} + u_{1j}
 \end{aligned}
 \tag{D3}$$

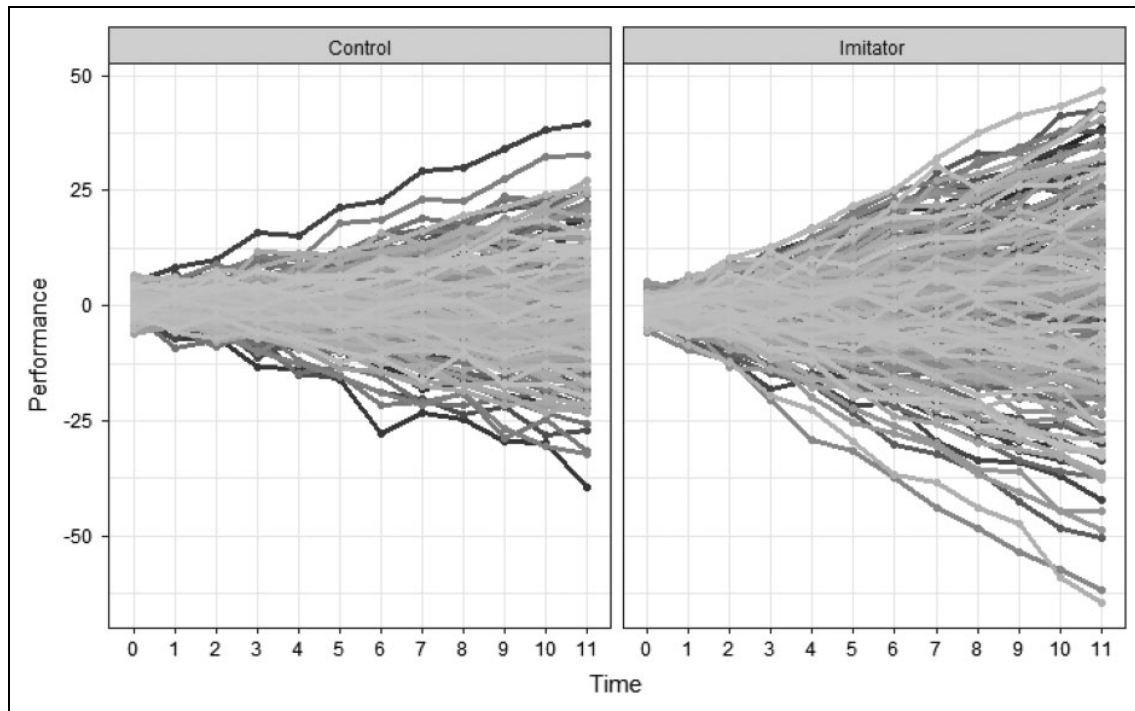


Figure D2. Spaghetti plot for longitudinal between firm performance variability example. Control firms are displayed on the left and imitator firms are displayed on the right.

The covariance matrices for the random effects were allowed to vary across the groups; see the covariance matrices presented below. The superscripts are used to denote group membership.

$$\begin{pmatrix} u_{0j}^0 \\ u_{1j}^0 \end{pmatrix} = \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0^0}^2 & \sigma_{u_0^0 u_1^0} \\ \sigma_{u_0^0 u_1^0} & \sigma_{u_1^0}^2 \end{bmatrix} \right) \\ \begin{pmatrix} u_{0j}^1 \\ u_{1j}^1 \end{pmatrix} = \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0^1}^2 & \sigma_{u_0^1 u_1^1} \\ \sigma_{u_0^1 u_1^1} & \sigma_{u_1^1}^2 \end{bmatrix} \right) \quad (\text{D4})$$

Appendix E

Alternative Models

The analyses described here are not meant to be a tutorial on a fixed modeling process. In other words, this is not a step by step tutorial. The research to provide such guidance in an exploratory model discovery process has not been conducted. This tutorial assumes a confirmatory modeling process where a model of theoretical interest is tested using the mixed-effects location-scale model. Traditional multilevel models are also presented to provide the reader with a familiar context and to allow us to more clearly illustrate the new research questions that can be answered using the mixed-effects location-scale model. The models are presented generically for both longitudinal and nested cross-sectional designs without reference to specific outcome variables.

The discussion proceeds from least complex to more complex starting with an empty general linear model and building to the mixed-effects location-scale model that has random intercepts and slopes on the location and scale sides of the model. Differences between the successive models are

denoted by bold font. In the covariance matrices, only parameters below the main diagonal are bolded because those above the diagonal are identical. R code to estimate these models is available in the online supplemental material.

The empty model contains no predictors on either side of the model; thus, this model simply provides an estimate of the overall population mean and the amount of variability around that mean. This is the model against which a random intercept multilevel model is typically compared.

Empty Model

	Location	Scale
Level 1	$Y_{ij} = B_{0j} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j}$
Level 2	$B_{0j} = \gamma_{00}$	$\alpha_{0j} = \tau_{00}$
Combined	$Y_{ij} = \gamma_{00} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \tau_{00}$

Random Intercept MLM

The random intercept multilevel model is the model that is typically used to calculate an intraclass correlation that measures the proportion of Level 2 variability or provides an estimate of the average covariance between the Level 1 units. In nested cross-sectional data (e.g., where employees are nested within a team), this would represent the proportion of variability that is between groups and estimate the average amount of covariance between individuals in the same group. A similar interpretation is given for longitudinal data where time is viewed as being nested within person/firm. The ICC provides an estimate of the amount of between person/firm variability and an estimate of the average amount of covariance between time points from the same person/firm. In terms of parameters estimated, the only different between the random intercept model and the empty model is the addition of the random intercept variance.

	Location	Scale
Level 1	$Y_{ij} = B_{0j} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j}$
Level 2	$B_{0j} = \gamma_{00} + u_{0j}$	$\alpha_{0j} = \tau_{00}$
Combined	$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \tau_{00}$

$$u_{0j} \sim N(0, \sigma_{u_0}^2)$$

Random Intercepts for the Location and Scale

In the models considered so far, the amount of variability between and within groups has been considered constant. The first mixed-effects location-scale model that we estimate allows us to assess whether that is a reasonable assumption. This model allows us to answer a question that cannot be answered using a traditional multilevel model (i.e., do some Level 2 units have more within-group/person variability than others). Also, the magnitude of the covariance between the random intercepts can be evaluated to assess whether Level 2 units with high/low mean tend to be more more/less variable. The covariance matrix of the Level 2 random effects is presented in the equation below. The $\sigma_{u_0 v_0}$ term indicates that the random intercept terms are allowed to covary with one another. Once again, this model could apply to both longitudinal or nested cross-sectional data;

measurements could be obtained from individuals nested within groups or repeated measurements over time from individuals or companies.

	Location	Scale
Level 1	$Y_{ij} = B_{0j} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j}$
Level 2	$B_{0j} = \gamma_{00} + u_{0j}$	$\alpha_{0j} = \tau_{00} + v_{0j}$
Combined	$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}$	$\log(\sigma_{e_{ij}}) = \tau_{00} + v_{0j}$

$$\begin{pmatrix} u_{0j} \\ v_{0j} \end{pmatrix} \sim \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 v_0} \\ \sigma_{u_0 v_0} & \sigma_{v_0}^2 \end{bmatrix} \right)$$

Random Slope Location and Random Intercept for the Scale

In addition to random intercepts, random slopes can be included on both the location side and the scale side of the model. This indicates that the strength of the relationship between the outcome and the predictor depends upon group membership. A random slope allows both the fixed and random effect of the variable (see the bolded terms in the equation provided below). On the mean side, a random slope indicates that the magnitude of relationship between the Level 1 predictor and the mean level varies depending upon the group. For some groups there is a strong relationship such that those individuals with relatively higher Level 1 scores are predicted to have higher/lower mean outcome ratings. The relationship is not as strong for other groups. The covariance between the random slope and the random intercepts can also be investigated which would provide the researcher with information regarding whether those groups with higher/lower within group variability or means tend to have larger/smaller Level 1 slopes. As has been the case throughout this discussion, this model can be estimated using longitudinal or nested cross-sectional data.

	Location	Scale
Level 1	$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \bar{X}_{\cdot j}) + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j}$
Level 2	$B_{0j} = \gamma_{00} + u_{0j}$ $B_{1j} = \gamma_{10} + u_{1j}$	$\alpha_{0j} = \tau_{00} + v_{0j}$
Combined	$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_{\cdot j}) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_{\cdot j}) + e_{ij}$	$\log(\sigma_{e_{ij}}) = \tau_{00} + v_{0j}$

$$\begin{pmatrix} u_{0j} \\ v_{0j} \\ u_{1j} \end{pmatrix} = \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 v_0} & \sigma_{u_0 u_1} \\ \sigma_{v_0 u_0} & \sigma_{v_0}^2 & \sigma_{v_0 u_1} \\ \sigma_{u_1 u_0} & \sigma_{u_1 v_0} & \sigma_{u_1}^2 \end{bmatrix} \right)$$

Random Slopes on Both the Location Side and the Scale Side of the Model

The final variance partition is to allow similarity to have a random slope on the scale side of the model. This parameter has a similar interpretation as the random slope on the mean side of the model. Specifically, the effect of the Level 1 predictor on variability depends on the group. For some groups, the variability is decreased/increased as the relative Level 1 variable changes while for others it does not. Once again, the covariance among the random effects can be interpreted as well. For example, the covariance among the random slopes assesses whether those groups with stronger/

weaker Level 1 mean effects have stronger/weaker effects on variability. This model is depicted in Figure 4 in the article.

	Location	Scale
Level 1	$Y_{ij} = B_{0j} + B_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j} + \alpha_{1j}(X_{ij} - \bar{X}_j)$
Level 2	$B_{0j} = \gamma_{00} + u_{0j}$ $B_{1j} = \gamma_{10} + u_{1j}$	$\alpha_{0j} = \tau_{00} + v_{0j}$ $\alpha_{1j} = \tau_{10} + v_{1j}$
Combined	$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_j) + u_{0j} + u_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$	$\log(\sigma_{e_{ij}}) = \tau_{00} + \tau_{10}(X_{ij} - \bar{X}_j) + v_{0j} + v_{1j}(X_{ij} - \bar{X}_j)$

$$\begin{pmatrix} u_{0j} \\ v_{0j} \\ u_{1j} \\ v_{1j} \end{pmatrix} \sim \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 v_0} & \sigma_{u_0 u_1} & \sigma_{u_0 v_1} \\ \sigma_{v_0 u_0} & \sigma_{v_0}^2 & \sigma_{v_0 u_1} & \sigma_{v_0 v_1} \\ \sigma_{u_1 u_0} & \sigma_{u_1 v_0} & \sigma_{u_1}^2 & \sigma_{u_1 v_1} \\ \sigma_{v_1 u_0} & \sigma_{v_1 v_0} & \sigma_{v_1 u_1} & \sigma_{v_1}^2 \end{bmatrix} \right)$$

Predicting the Variance Components

Once the variance partitioning is completed, predictors of interest can be added to explain the various sources of variation. For example, Level 2 predictors can be used to explain any of the four random components that have been discussed (i.e., random intercepts and slopes on both the mean side and the variability side of the model). Although this is more complex than a traditional multi-level model, explaining the variability components is conceptually the same. The leader-member exchange example presented in the article is an example of this sort of model.

	Location	Scale
Level 1	$Y_{ij} = B_{0j} + B_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$	$\log(\sigma_{e_{ij}}) = \alpha_{0j} + \alpha_{1j}(X_{ij} - \bar{X}_j)$
Level 2	$B_{0j} = \gamma_{00} + \gamma_{01}(\bar{X}_j - \bar{X}_{..}) + \gamma_{02}W_j + u_{0j}$ $B_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$	$\alpha_{0j} = \tau_{00} + \tau_{01}(\bar{X}_j - \bar{X}_{..}) + \tau_{02}W_j + v_{0j}$ $\alpha_{1j} = \tau_{10} + \tau_{11}W_j + v_{1j}$ $\log(\sigma_{e_{ij}}) = \tau_{00} + \tau_{10}(X_{ij} - \bar{X}_j) + \tau_{01}(\bar{X}_j - \bar{X}_{..}) + \tau_{02}W_j + v_{0j} + \tau_{11}W_j(X_{ij} - \bar{X}_j) + v_{1j}(X_{ij} - \bar{X}_j)$
Combined	$Y_{ij} = \gamma_{00} + \gamma_{10}(X_{ij} - \bar{X}_j) + \gamma_{01}(\bar{X}_j - \bar{X}_{..}) + \gamma_{02}W_j + u_{0j} + \gamma_{11}W_j(X_{ij} - \bar{X}_j) + u_{1j}(X_{ij} - \bar{X}_j) + e_{ij}$	

$$\begin{pmatrix} u_{0j} \\ v_{0j} \\ u_{1j} \\ v_{1j} \end{pmatrix} \sim \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{u_0}^2 & \sigma_{u_0 v_0} & \sigma_{u_0 u_1} & \sigma_{u_0 v_1} \\ \sigma_{v_0 u_0} & \sigma_{v_0}^2 & \sigma_{v_0 u_1} & \sigma_{v_0 v_1} \\ \sigma_{u_1 u_0} & \sigma_{u_1 v_0} & \sigma_{u_1}^2 & \sigma_{u_1 v_1} \\ \sigma_{v_1 u_0} & \sigma_{v_1 v_0} & \sigma_{v_1 u_1} & \sigma_{v_1}^2 \end{bmatrix} \right)$$

In summary, mixed-effects location-scale model allows researchers to answer research questions regarding heterogeneous variances that cannot be addressed using traditional multilevel models. The

mixed-effects location-scale model allows researchers to quantify and predict heterogeneous residual variances while simultaneously accounting for any mean predictors of interest. Level 1 or Level 2 variables may be utilized to predict heterogeneous variance, and Level 1 predictors may be allowed to have effects that vary across groups. Figures 2 to 4 provide depictions that highlight the differences between the traditional multilevel models and those discussed here.

Appendix F

Calculations for Effect Size

The scale side of the MELS/heterogeneous variance model uses a log transformation. Thus, the outcome of the log-linear model is the log standard deviation. This is unlikely to be intuitive in terms of interpreting effect sizes; however, transformations can be performed to make the metric more interpretable. Exponentiating both the predictor and outcome sides of the equation puts the outcome on the standard deviation metric. To demonstrate how this works we will thoroughly explain the effect of a one standard deviation increase in transformational leadership when similarity is at its mean of zero. Similarity_{ij} and $\overline{\text{Similarity}}_{\cdot j}$ (i.e., an individual having a score of zero where the group mean is also zero) is important for this demonstration simply because it allows us to examine the effect that transformational leadership has on the predicted standard deviation without considering the interaction term.

$$\log(\sigma_{eij}) = \alpha_{0j} + \alpha_{1j}(\text{Similarity}_{ij} - \overline{\text{Similarity}}_{\cdot j})$$

$$\alpha_{0j} = \tau_{00} + \tau_{01}(\overline{\text{Similarity}}_{\cdot j}) + \tau_{02}(\text{Transformational}_j - \overline{\text{Transformational}}_{\cdot}) + v_{0j}$$

$$\alpha_{1j} = \tau_{10} + \tau_{11}(\text{Transformational}_j - \overline{\text{Transformational}}_{\cdot}) + v_{1j}$$

Substituting the similarity values of interest into the equation presented below results in the simplified formula. Thus, assuming a constant v_{0j} at zero, the predicted log standard deviation for a group with a leader at the average level of transformational leadership equal is τ_{00} and the predicted value for a group with leader that is one standard deviation above the mean in transformational is $\tau_{00} + \tau_{02}$.

$$\log(\sigma_{eij}) = \tau_{00} + \tau_{02}(\text{Transformational}_j - \overline{\text{Transformational}}_{\cdot}) + v_{0j}$$

As state previously, converting this to the standard deviation scale from the log standard deviation scale is done by exponentiation. Thus, the group with a mean transformational leadership score is predicted to have a standard deviation of $e^{\tau_{00}} = e^{-.14} = .87$. The same calculation results in $e^{\tau_{00} + \tau_{02}} = e^{-.14 - .07} = .81$ for a group with leader that is one unit above the mean on transformational leadership. The percentage change is calculated as follows: $[(.81 - .87)/.87] \times 100 \approx 7\%$.

Authors' Note

We would like to thank Justin Jones as well as special feature editor Rory Eckardt, the other editors, and anonymous reviewers for the Feature Topic on New Approaches to Multilevel Methods and Statistics for their comments and suggestions throughout the review of this article.

Houston F. Lester is also affiliated with Center for Innovations in Quality, Effectiveness and Safety, Michael E. DeBakey VA Medical Center, Houston, Texas.


Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was partially funded by the first author's VA Health Services Research & Development (HSR&D) Postdoctoral Fellowship: HX-17-014.

ORCID iD

Houston F. Lester  <https://orcid.org/0000-0003-3961-6219>

Notes

1. Please note that “group” could refer to any Level 2 grouping variable. In the first illustrative example this is people nested within a supervisor's workgroup, and in the second illustrative example this is time nested within firm. However, the grouping variable could represent individuals, teams, departments, organizations, industries, and so forth.
2. This article is not concerned with contextual effects.

References

- Aitkin, M. (1987). Modeling variance heterogeneity in normal regression using GLIM. *Applied Statistics*, 36(3), 332-339. doi:10.2307/2347792
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of Hirotugu Akaike* (pp. 199-213). New York, NY: Springer. doi.org/10.1007/978-1-4612-1694-0_15
- Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling*, 25(3), 359-388. doi:10.1080/10705511.2017.1406803
- Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis using Mplus: Technical implementation* (Tech. Rep. 3). Retrieved from www.statmodel.com/download/Bayes3.pdf
- Asparouhov, T., & Muthén, B. (2016). General random effect latent variable modeling: Random subjects, items, contexts, and parameters. In J. R. Harring, L. M. Stapleton, & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research: Addressing practical issues found in real-world applications* (pp. 163-192). Charlotte, NC: Information Age.
- Balasubramanian, N., & Lieberman, M. B. (2010). Industry learning environments and the heterogeneity of firm performance. *Strategic Management Journal*, 31(4), 390-412. doi:10.1002/smj.816
- Barnes, C. M., Reb, J., & Ang, D. (2012). More than just the mean: Moving to a dynamic view of performance-based compensation. *Journal of Applied Psychology*, 97(3), 711-718. doi:10.1037/a0026927
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. doi:10.18637/jss.v067.i01
- Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical Transactions of the Royal Society of London*, 53, 370-418.
- Bidwell, M., & Keller, J. R. (2014). Within or without? How firms combine internal and external labor markets to fill jobs. *Academy of Management Journal*, 57(4), 1035-1055. doi:10.5465/amj.2012.0119
- Bliese, P. D., & Ployhart, R. E. (2002). Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods*, 5(4), 362-387. doi:10.1177/109442802237116
- Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford.
- Browne, W. J. (2017). *MCMC Estimation in MLwiN v3.00*. University of Bristol, Centre for Multilevel Modelling.
- Bryk, A. S., & Raudenbush, S. W. (1988). Heterogeneity of variance in experimental studies: A challenge to conventional interpretations. *Psychological Bulletin*, 104(3), 396-404. doi:10.1037//0033-2909.104.3.396

- Bürkner, P. C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80. doi:10.18637/jss.v080.i01
- Carnahan, S., Agarwal, R., & Campbell, B. A. (2012). Heterogeneity in turnover: The effect of relative compensation dispersion of firms on the mobility and entrepreneurship of extreme performers. *Strategic Management Journal*, 33(12), 1411-1430. doi:10.1002/smj.1991
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., . . . Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76. doi:10.18637/jss.v076.i01
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Wadsworth.
- Cen, L., Lu, H., & Yang, L. (2013). Investor sentiment, disagreement, and the breadth–return relationship. *Management Science*, 59(5), 1076-1091. doi:10.1287/mnsc.1120.1633
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis. *Journal of Applied Psychology*, 83(2), 234-246. doi:10.1037//0021-9010.83.2.234
- Chin, M. K., & Semadeni, M. (2017). CEO political ideologies and pay egalitarianism within top management teams. *Strategic Management Journal*, 38(8), 1608-1625. doi:10.1002/smj.2608
- Chordia, T., Subrahmanyam, A., & Anshuman, V. R. (2001). Trading activity and expected stock returns. *Journal of Financial Economics*, 59(1), 3-32. doi:10.1016/S0304-405X(00)00080-5
- Chowdhury, S. K., & Endres, M. L. (2010). The impact of client variability on nurses' occupational strain and injury: Cross-level moderation by safety climate. *Academy of Management Journal*, 53(1), 182-198. doi:10.5465/amj.2010.48037720
- Chun, J. S., & Choi, J. N. (2014). Members' needs, intragroup conflict, and group performance. *Journal of Applied Psychology*, 99(3), 437-450. doi:10.1037/a0036363
- Cobb, J. A., & Stevens, F. G. (2017). These unequal states: Corporate organization and income inequality in the United States. *Administrative Science Quarterly*, 62(2), 304-340. doi:10.1177/0001839216673823
- Culpepper, S. A. (2010). Studying individual differences in predictability with gamma regression and nonlinear multilevel models. *Multivariate Behavioral Research*, 45(1), 153-185.
- Dalal, R. S., Bhawe, D. P., & Fiset, J. (2014). Within-person variability in job performance: A theoretical review and research agenda. *Journal of Management*, 40(5), 1396-1436. doi:10.1177/0149206314532691
- Dalal, R. S., Meyer, R. D., Bradshaw, R. P., Green, J. P., Kelly, E. D., & Zhu, M. (2015). Personality strength and situational influences on behavior: A conceptual review and research agenda. *Journal of Management*, 41(1), 261-287. doi:10.1177/0149206314557524
- Davidian, M., & Giltinan, D. (1995). *Nonlinear models for repeated measurement data*. Monographs on Statistics and Applied Probability. London, UK: Chapman & Hall.
- Depaoli, S., & van de Schoot, R. (2015). Improving transparency and replication in Bayesian statistics: The WAMBS-checklist. *Psychological Methods*, 22(2), 240-261. doi:10.1037/met0000065
- DeRue, D. S., Hollenbeck, J., Ilgen, D., & Feltz, D. (2010). Efficacy dispersion in teams: Moving beyond agreement and aggregation. *Personnel Psychology*, 63(1), 1-40. doi:10.1111/j.1744-6570.2009.01161.x
- Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., & Ferris, G. R. (2012). A meta-analysis of antecedents and consequences of leader-member exchange: Integrating the past with an eye toward the future. *Journal of Management*, 38(6), 1715-1759. doi:10.1177/0149206311415280
- Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2), 121-138. doi:10.1037/1082-989x.12.2.121
- Erdogan, B., & Bauer, T. N. (2010). Differentiated leader–member exchanges: The buffering role of justice climate. *Journal of Applied Psychology*, 95(6), 1104-1120. doi:10.1037/a0020578
- Figueiredo, P. N. (2011). The role of dual embeddedness in the innovative performance of MNE subsidiaries: Evidence from Brazil. *Journal of Management Studies*, 48(2), 417-440. doi:10.1111/j.1467-6486.2010.00965.x
- Fredrickson, J. W., Davis-Blake, A., & Sander, W. G. (2010). Sharing the wealth: Social comparisons and pay dispersion in the CEO's top team. *Strategic Management Journal*, 31(10), 1031-1053. doi:10.1002/smj.848

- Gabriel, A. S., & Diefendorff, J. M. (2015). Emotional labor dynamics: A momentary approach. *Academy of Management Journal*, 58(6), 1804-1825. doi:10.5465/amj.2013.1135
- Gabriel, A. S., Diefendorff, J. M., Chandler, M. M., Moran, C. M., & Greguras, G. J. (2014). The dynamic relationships of work affect and job satisfaction with perceptions of fit. *Personnel Psychology*, 67(2), 389-420. doi:10.1111/peps.12042
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.
- Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6), 997-1016. doi:10.1007/s11222-013-9416-2
- Goldstein, H. (2011). *Multilevel statistical models* (Vol. 922). John Wiley.
- Gonzalez-Mulé, E., DeGeest, D. S., McCormick, B. W., Seong, J. Y., & Brown, K. G. (2014). Can we get some cooperation around here? The mediating role of group norms on the relationship between team personality and individual helping behaviors. *Journal of Applied Psychology*, 99(5), 988-999. doi:10.1037/a0037278
- González-Romá, V., Peiró, J. M., & Tordera, N. (2002). An examination of the antecedents and moderator influences of climate strength. *Journal of Applied Psychology*, 87(3), 465-473. doi:10.1037//0021-9010.87.3.465
- Gooty, J., & Yammarino, F. J. (2016). The leader-member exchange relationship: A multisource, cross-level investigation. *Journal of Management*, 42(4), 915-935. doi:10.1177/0149206313503009
- Greenland, S. G. (2006). Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International Journal of Epidemiology*, 35(3), 765-775. doi:10.1093/ije/dyi312
- Guo, G. (2017). Demystifying variance in performance: A longitudinal multilevel perspective. *Strategic Management Journal*, 38(6), 1327-1342. doi:10.1002/smj.2555
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44(3), 461-466. doi:10.2307/1913974
- Hausknecht, J. P., & Holwerda, J. A. (2013). When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. *Organization Science*, 24(1), 210-225. doi:10.1287/orsc.1110.0720
- Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, 64(2), 627-634. doi:10.1111/j.1541-0420.2007.00924.x
- Hedeker, D., & Nordgren, R. (2013). MIXREGLS: A program for mixed-effects location scale analysis. *Journal of Statistical Software*, 52, 1-38. doi:10.18637/jss.v052.i12
- Henderson, D. J., Liden, R. C., Glibkowski, B. C., & Chaudhry, A. (2009). LMX differentiation: A multilevel review and examination of its antecedents and outcomes. *Leadership Quarterly*, 20(4), 517-534. doi:10.1016/j.leaqua.2009.04.003
- Hoffman, L. (2007). Multilevel models for examining individual differences in within-person variation and covariation over time. *Multivariate Behavioral Research*, 42(4), 609-629. doi:10.1080/00273170701710072
- Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593-1623.
- Hofmann, D. A. (2002). Issues in multilevel research: Theory development, measurement, and analysis. In S. G. Rogelberg (Ed.), *Handbook of research methods in industrial and organizational psychology* (pp. 247-274). Malden, MA: Blackwell.
- Huang, J. L., Ford, J. K., & Ryan, A. M. (2017). Ignored no more: Within-person variability enables better understanding of training transfer. *Personnel Psychology*, 70(3), 557-596. doi:10.1111/peps.12155
- Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review of Psychology*, 56, 517-543. doi:10.1146/annurev.psych.56.091103.070250
- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69(1), 85-98. doi:10.1037//0021-9010.69.1.85

- James, L. R., Demaree, R. G., & Wolf, G. (1993). rwg: An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78(2), 306-309. doi:10.1037//0021-9010.78.2.306
- Jaskiewicz, P., Block, J. H., Miller, D., & Combs, J. G. (2017). Founder versus family owners' impact on pay dispersion among non-CEO top managers: Implications for firm performance. *Journal of Management*, 43(5), 1524-1552. doi:10.1177/0149206314558487
- Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York, NY: Guilford.
- Kellermanns, F. W., Walter, J., Lechner, C., & Floyd, S. W. (2005). The lack of consensus about strategic consensus: Advancing theory and research. *Journal of Management*, 31(5), 719-737. doi:10.1177/0149206305279114
- Kim, J. Y., Finkelstein, S., & Halebian, J. (2015). All aspirations are not created equal: The differential effects of historical and social aspirations on acquisition behavior. *Academy of Management Journal*, 58(5), 1361-1388. doi:10.5465/amj.2012.1102
- Kreft, I. G., De Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, 30(1), 1-21. doi:10.1207/s15327906mbr3001_1
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, Jags, and Stan* (2nd ed.). New York, NY: Academic Press.
- Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The time has come: Bayesian methods for data analysis in the organizational sciences. *Organizational Research Methods*, 15(4), 722-752. doi:10.1177/1094428112457829
- Kuppens, T., & Yzerbyt, V. Y. (2014). Predicting variability: Using multilevel models to assess differences in variances. *European Journal of Social Psychology*, 44(7), 691-700. doi:10.1002/ejsp.2028
- Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied linear statistical models* (5th ed.). New York, NY: McGraw-Hill.
- Laamanen, T., & Keil, T. (2008). Performance of serial acquirers: Toward an acquisition program perspective. *Strategic Management Journal*, 29(6), 663-672. doi:10.1002/smj.670
- Lang, J. W. B., Bliese, P. D., & de Voogt, A. (2018). Modeling consensus emergence in groups using longitudinal multilevel methods. *Personnel Psychology*, 71(2), 255-281. doi:10.1111/peps.12260
- Leckie, G., French, R., Charlton, C., & Browne, W. (2014). Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics*, 39(5), 307-332. doi:10.3102/1076998614546494
- Lenox, M. J., Rockart, S. F., & Lewin, A. Y. (2010). Does interdependency affect firm and industry profitability? An empirical test. *Strategic Management Journal*, 31(2), 121-139. doi:10.1002/smj.811
- Leonard, T. (1975). A Bayesian approach to the linear model with unequal variances. *Techometrics*, 17(1), 95-105. doi:10.2307/1268006
- Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9), 1989-2001. doi:10.1016/j.jmva.2009.04.008
- Liden, R. C., Erdogan, B., Wayne, S. J., & Sparrowe, R. T. (2006). Leader-member exchange, differentiation, and task interdependence: Implications for individual and group performance. *Journal of Organizational Behavior*, 27(6), 723-746. doi:10.1002/job.409
- Liden, R. C., Wayne, S. J., & Stilwell, D. (1993). A longitudinal study on the early development of leader-member exchanges. *Journal of Applied Psychology*, 78, 662-674. doi:10.1037//0021-9010.78.4.662
- Lievens, F., Lang, J. W., De Fruyt, F., Corstjens, J., Van de Vijver, M., & Bledow, R. (2018). The predictive power of people's intraindividual variability across situations: Implementing whole trait theory in assessment. *Journal of Applied Psychology*, 103(7), 753-771. doi:10.1037/apl0000280
- Lievens, F., Sanchez, J. I., Bartram, D., & Brown, A. (2010). Lack of consensus among competency ratings of the same occupation: Noise or substance? *Journal of Applied Psychology*, 95(3), 562-571. doi:10.1037/a0018035

- Lim, E. (2018). Attainment discrepancy and new geographic market entry: The moderating roles of vertical pay disparity and horizontal pay dispersion. *Journal of Management Studies*. Advance online publication. doi: 10.1111/joms.12430
- Lin, X., Mermelstein, R. J., & Hedeker, D. (2018a). A shared parameter location scale mixed effect model for EMA data subject to informative missing. *Health Services and Outcomes Research Methodology*, 18(4), 227-243. doi:10.1007/s10742-018-0184-5
- Lin, X., Mermelstein, R. J., & Hedeker, D. (2018b). A three level Bayesian mixed effects location scale model with an application to ecological momentary assessment (EMA) data. *Statistics in Medicine*, 31(26), 2108-2119. doi:10.1002/sim.7627
- Lindley, D. V. (1971). The estimation of many parameters. In V. P. Godambe & D. A. Sprott (Eds.), *Foundations of statistical inference* (pp. 435-455). Toronto, Canada: Holt, Rinehart, and Winston.
- Lynch, S. M. (2007). *Introduction to applied bayesian statistics and estimation for social scientists*. New York: Springer.
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3), 356-376. doi:10.5465/amr.2001.4845785
- Martin, R., Thomas, G., Legood, A., & Dello Russo, S. (2018). Leader-member exchange (LMX) differentiation and work outcomes: Conceptual clarification and critical review. *Journal of Organizational Behavior*, 39(2), 151-168. doi:10.1002/job.2202
- Matta, F. K., Scott, B. A., Colquitt, J. A., Koopman, J., & Passantino, L. G. (2017). Is consistently unfair better than sporadically fair? *An investigation of justice variability and stress*. *Academy of Management Journal*, 60(2), 743-770. doi:10.5465/amj.2014.0455
- McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan (Vol. 122)*. Boca Raton, FL: CRC Press.
- Melkonyan, T., & Safra, Z. (2016). Intrinsic variability in group and individual decision making. *Management Science*, 62(9), 2651-2667. doi:10.1287/mnsc.2015.2255
- Messersmith, J. G., Guthrie, J. P., Ji, Y. Y., & Lee, J. Y. (2011). Executive turnover: The influence of dispersion and other pay system characteristics. *Journal of Applied Psychology*, 96(3), 457-469. doi:10.1037/a0021654
- Minbashian, A., & Luppino, D. (2014). Short-term and long-term within-person variability in performance: An integrative model. *Journal of Applied Psychology*, 99(5), 898-914. doi:10.1037/a0037402
- Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *Journal of Applied Psychology*, 95(5), 793-806.
- Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.
- Neal, R. M. (2010). MCMC using hamiltonian dynamics. In S. Brooks, A. Gelman, G. L. Jones, & X.-L. Meng (Eds.), *The handbook of Markov Chain Monte Carlo* (pp. 113-162). Boca Raton, FL: Chapman & Hall/CRC Press.
- Pak, J., & Kim, S. (2018). Team manager's implementation, high performance work systems intensity, and performance: A multilevel investigation. *Journal of Management*, 44(7), 2690-2715. doi:10.1177/0149206316646829
- Patel, P. C., Kohtamäki, M., Parid, V., & Wincent, J. (2015). Entrepreneurial orientation-as-experimentation and firm performance: The enabling role of absorptive capacity. *Strategic Management Journal*, 36, 1739-1749. doi:10.1002/smj.2310
- Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Oxford, UK: Oxford University Press.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects in S and S-Plus*. New York, NY: Springer.
- Ployhart, R. E., Weekley, J. A., & Baughman, K. (2006). The structure and function of human capital emergence: A multilevel examination of the attraction-selection-attrition model. *Academy of Management Journal*, 49(4), 661-677. doi:10.5465/amj.2006.22083023
- Plummer, M. (2013). JAGS: Just another gibbs sampler. Retrieved from <http://mcmc-jags.sourceforge.net/>

- Posen, H. E., & Martignoni, D. (2018). Revisiting the imitation assumption: Why imitation may increase, rather than decrease, performance heterogeneity. *Strategic Management Journal*, 39(5), 1350-1369. doi:10.1002/smj.2751
- Pugach, O., Hedeker, D., Richmond, M. J., Sokolovsky, A., & Mermelstein, R. (2014). A bivariate mixed-effects location-scale model with application to ecological momentary assessment (EMA) data. *Health Services and Outcomes Research Methodology*, 14(4), 194-212. doi:10.1007/s10742-014-0126-9
- Raudenbush, S. W. (1988). Estimating change in dispersion. *Journal of Educational Statistics*, 13(2), 148-171. doi:10.3102/10769986013002148
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods (Vol. 1)*. Thousand Oaks, CA: Sage.
- Reb, J., & Greguras, G. J. (2010). Understanding performance ratings: Dynamic performance, attributions, and rating purpose. *Journal of Applied Psychology*, 95(1), 213-220. doi:10.1037/a0017237
- Roberson, Q. M., & Williamson, I. O. (2012). Justice in self-managing teams: The role of social networks in the emergence of procedural justice climates. *Academy of Management Journal*, 55(3), 685-701. doi:10.5465/amj.2009.0491
- Schyns, B. (2006). Are group consensus in leader-member exchange (LMX) and shared work values related to organizational outcomes? *Small Group Research*, 37(1), 20-35. doi:10.1177/1046496405281770
- Scott, B. A., Barnes, C. M., & Wagner, D. T. (2012). Chameleonic or consistent? A multilevel investigation of emotional labor variability and self-monitoring. *Academy of Management Journal*, 55, 905-926. doi:10.5465/amj.2010.1050
- Sitzmann, T., Ployhart, R. E., & Kim, Y. (2019). A process model linking occupational strength to attitudes and behaviors: The explanatory role of occupational personality heterogeneity. *Journal of Applied Psychology*, 104, 247-269. doi:10.1037/apl0000352
- Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2007). *OpenBUGS user manual, version 3.0.2*. Cambridge, UK: MRC Biostatistics Unit.
- Sriram, S., Chintagunta, P. K., & Manchanda, P. (2015). Service quality variability and termination behavior. *Management Science*, 61, 2739-2759. doi:10.1287/mnsc.2014.2105
- Trevor, C. O., Reilly, G., & Gerhart, B. (2012). Reconsidering pay dispersion's effect on the performance of interdependent work: Reconciling sorting and pay inequality. *Academy of Management Journal*, 55, 585-610. doi:10.5465/amj.2006.0127
- Wales, W. J., Patel, P. C., & Lumpkin, G. T. (2013). In pursuit of greatness: CEO narcissism, entrepreneurial orientation, and firm performance variance. *Journal of Management Studies*, 50(6), 1041-1069. doi:10.1111/joms.12034
- Walters, R. W., Hoffman, L., & Templin, J. (2018). The power to detect and predict individual differences in intra-individual variability using the mixed-effects location-scale model. *Multivariate Behavioral Research*, 53(3), 360-374. doi:10.1080/00273171.2018.1449628
- Wang, H., & Choi, J. (2013). A new look at the corporate social-financial performance relationship: The moderating roles of temporal and interdomain consistency in corporate social performance. *Journal of Management*, 39(2), 416-441. doi:10.1177/0149206310375850
- Watts, A., Walters, R. W., Hoffman, L., & Templin, J. (2016). Intra-individual variability of physical activity in older adults with and without mild Alzheimer's disease. *PLOS ONE*, 11(4). doi:10.1371/journal.pone.0153898

Author Biographies

Houston F. Lester is an instructor of medicine at the Baylor College of Medicine and an investigator at the Michael E. DeBakey VA Medical Center, Houston, Texas. He received a PhD in Educational Psychology with a specialization in Quantitative, Qualitative, and Psychometric Methods from the University of Nebraska-Lincoln.

His research focuses on mixed-effects location-scale models and team psychometrics. His research has been funded by the Agency for Healthcare Research and Quality.

Kristin L. Cullen-Lester is an assistant professor at the C. T. Bauer College of Business at the University of Houston. She received a PhD in Industrial-Organizational Psychology from Auburn University. Her research examines relational aspects of leadership, including leaders' social networks and the role of organizational networks in shared leadership, complex collaboration, and the implementation of strategic and large-scale change. Her research has been funded by the National Science Foundation.

Ryan W. Walters is an assistant professor at the Creighton University Department of Medicine. He received a PhD in Quantitative Psychology from the University of Nebraska-Lincoln. His research focuses on multilevel modeling, mixed-effects location-scale models, and applying statistics in medicine. His research has been funded by the Robert Wood Johnson Foundation among other national funding agencies.